

Duke Teaching and Leading EBP

A Workshop for Educators and Champions of Evidence-Based Practice



Table of Contents

Table of Contents

Intro to Evidence-Based Practice.....	5
Core Concepts on Teaching EBP.....	5
Evidence Cycle	6
Article with an Example of Applying Evidence	10
Clinical Question	11
The Clinical Question and Information Resources	11
PICO Table Handout	12
Searching for Evidence.....	13
Formulating Search Strategy	13
EBM Resources Comparison Chart.....	14
PubMed Basics	17
Saving Searches in PubMed: My NCBI.....	20
Setting Up Filters in PubMed: My NCBI.....	21
Therapy.....	22
Therapy Core Concepts	22
Therapy Critical Appraisal Form.....	23
Therapy Critical Appraisal Form, Non-inferiority trials.....	24
FRISBE Therapy Critical Appraisal Worksheet with Key Learning Points	26
Therapy Formulas	29
Therapy Example Exercise	30
Therapy Example Exercise with ANSWERS	31
Harm	33
Harm Core Concepts.....	33
Harm Critical Appraisal Form, Cohort Study	34
Harm Critical Appraisal Form, Case-Control Study.....	35
Harm Sample Packet: The Class Study (RCT)	36
ANSWERS to HARM/RCT Example Teaching Package	45
Harm Sample Packet #2: Post-Op Delirium - Case-Control Study.....	52
ANSWERS CRITICAL REVIEW FORM FOR HARM – Example #2 - Case-Control Study	54
Harm Sample Packet #3: Case Series/Report.....	56
ANSWERS for Case Series/Report Teaching Harm/Etiology Example #3.....	58

Prognosis.....	61
Prognosis Core Concepts	61
Prognosis Critical Appraisal Form.....	62
Prognosis Teaching Package #1: Hepatitis A- A deadly partner?.....	63
PROGNOSIS Teaching Package: Hep C Critical Appraisal ANSWERS	64
Prognosis Example #3	65
Prognosis Example Coffee ANSWERS	68
Critical Review Form for Prognosis ANSWERS	70
Diagnosis.....	73
Diagnosis Core Concepts.....	73
Diagnostic Test Critical Appraisal Form	74
Differential Diagnosis Critical Appraisal Form.....	75
Screening Critical Appraisal Form.....	76
Diagnostic Test Worksheets.....	77
Diagnosis Teaching Package Example	80
Diagnosis Teaching Package ANSWERS and Teaching Tips.....	86
Summary Methods SR MA Guidelines	93
Systematic Review Core Concepts.....	93
Systematic Review.....	94
Comparing different types of reviews.....	95
Systematic Review Class Worksheets	97
Systematic Review Example of Teaching Package ANSWERS (Therapy Unstable Angina)	101
Qualitative and Other Methods.....	104
Other Topics Core Concepts.....	104
Quantitative vs. Qualitative Methods Comparison	105
Qualitative Methods Appraisal Form	106
Clinical Decision Analysis Appraisal Form.....	107
Clinical Practice Guidelines Appraisal Form	108
Economic Analysis Appraisal Form	109
Qualitative Example Teaching Package.....	110
Qualitative Research Example ANSWERS	116
Teaching Strategies.....	120
Teaching Strategies Introduction: Teaching Tips and Materials	120
Tips for Creating Examples: Writer Suggestions to Consider	121
Ridicularum Exemplario.....	123
Filling the Tool Bag Strategies for making it fun and effective	124

Feedback: 6T's Teaching Tips.....	127
Curriculum Planner Workbook.....	128
Evidence Based Clinical Practice Competency Grid	134
Sample Evidence-Based Medicine Curriculum.....	135
How to choose critical appraisal worksheet	137
How to use the rational clinical examination education guides.....	139
Educational Prescription: Duke University Medical Center.....	143
Appendix.....	144
Critical Appraisal Forms	144
Clinical Decision Analysis	144
Clinical Practice Guidelines.....	145
Diagnostic Test.....	146
Differential Diagnosis.....	147
Economic Analysis.....	148
Harm, Cohort Study.....	149
Harm, Case-Control Study	150
Prognosis.....	151
Qualitative Methods.....	152
Screening.....	153
Systematic Review.....	154
Therapy	155
Therapy, Non-inferiority trials.....	156
Teaching Papers	158
Diagnosis	158
Harm.....	180
Other.....	191
Prevention.....	194
Prognosis.....	198
Syntheses - SR, MA, CPG	211
Therapy.....	236

Intro to Evidence-Based Practice

Core Concepts on Teaching EBP

Notes on learning and teaching about **EBP Framework**:

1. **Background**: There are many concepts and key terms that have relevance across many of the specific content areas such as therapy, diagnosis, harm, and prognosis.
2. Key concepts and terms
 - Principles of EBP: **hierarchy of evidence** and role of patient **values and preferences**
 - The 5 A's of the evidence cycle
3. Clinical questions and searching
 - Clinical question formation: PICOTT (**p**atient, **i**ntervention, **c**omparison, **o**utcome, **t**ype of question and **t**ype of study you seek to find)
 - Sources of best evidence including pre-appraised resources, and searching the biomedical literature (including PubMed)
4. General principles pertaining to understanding data and assessing bias
 - Critical appraisal process including evaluating studies for risk of bias, understanding results and applying evidence to particular patients and populations
 - Random Error (chance) versus Systematic Error (bias)
 - Clinical versus statistical significance
 - Relationship between sample size and number of outcomes and confidence in estimates

Evidence Cycle

The five A's of the Evidence Cycle

Incorporating the best evidence into clinical care requires a systematic approach in order to be manageable. A clear series of steps known as the Evidence Cycle can provide an excellent paradigm to guide you through this process. The foundation of evidence-based care remains an excellent clinical evaluation. The clinician must **ASSESS** the patient and the problem to determine the pertinent issues, which may include a differential diagnosis, treatment decisions, or prognosis. The clinician must draw from this evaluation and **ASK** a clear, answerable question to be pursued. The next step is to efficiently **ACQUIRE** the evidence from an appropriate source. Potential sources include original research studies, systematic reviews, evidence-based journal abstracts, textbooks, and computerized decision support systems. With a potential source in hand, the clinician must **APPRAISE** the evidence to further examine its worth and reliability. Finally, the process must conclude by returning to the individual patient, as the clinician has to decide whether it is appropriate to **APPLY** the evidence to the particular patient and their unique values and circumstances. Evidence alone is never sufficient to direct decision making. Rather, it must be put into context with a patient's values.

ASSESS: Clinical Evaluation

The method of evidence-based clinical practice (EBCP) begins with a thoughtful assessment by a clinician who incorporates all the pertinent data. A common fallacy is that EBCP somehow devalues the fundamental tenets of the practice of medicine, specifically clinical expertise. A comprehensive understanding of pathophysiology and the thorough history and physical remain a critical starting point for the process.

ASK: Clinical Question Development

The first critical step is to clarify one or two key issues that come up in the course of caring for your patient and to develop a focused clinical question. Despite its critical place at the start of the evidence cycle, question development is often not a focus of training. In a survey of 417 internal medicine program directors, only 44% of programs with evidence-based medicine curricula included posing a focused question as an objective.ⁱ However, without this critical first step, the rest of the steps are immaterial.

The Anatomy of the Clinical Question (PICOTT)

One useful approach to framing a clinical question involves distilling the question into several key elements. In this framework, there are 4 components to every clinical question, the Patient population, Intervention, Comparison and Outcome (PICO).ⁱⁱ In addition, we add two "T"s to capture the type of question being asked (e.g. Therapy) and the type of study you would want to find (e.g. RCT). We can use this framework to clarify the steps that we must take to find the evidence we seek.

ACQUIRE: Searching for the Evidence

Armed with our well-built clinical questions, our attention next turns to finding the evidence in the medical literature. Many resources are currently available; therefore, we must learn to appreciate the pros and cons of each type to determine when each one can best be applied. We also have to learn how to access resources that can maximize our efficiency, such as a systematic review, clinical practice guideline or an evidence-based journal abstract.

APPRAISE: Critical Appraisal of the Evidence

Much of the initial attention in the realm of evidence-based medicine focused on the critical appraisal portion of the evidence cycle. A growing body of resources exists in various print and electronic formats to aid readers of the medical literature in the critical appraisal process. The following tables were abstracted from the *Users' Guides to the Medical Literature* from the evidence-based medicine working group. (See Table)

APPLY: Applying Evidence to the Patient

Every management decision requires value-laden deliberation and judgment. Each piece of evidence that we review adds something to our understanding of our patient's situation. However, we need to consider how to generalize the results from clinical trials to our individual patient. We must consider whether the patient populations and treatments or interventions are comparable to our setting. The final challenge is to combine the evidence and clinical expertise with compassion and patient values. Clinicians trying to engage the medical literature for best care must take the information from these studies to try to help individuals within the context of their own values and preferences.

How serious is the risk of bias?

Table extracted from *User's Guide to the Medical Literature, Evidence-Based Medicine Working Group*

(Note: Bold Text indicates the questions that can serve as your first screen for validity)

Therapy or prevention	<ul style="list-style-type: none"> • Was the assignment of patients to treatments randomized?
	<ul style="list-style-type: none"> • Were all of the patients who entered the trial properly accounted for and attributed at its conclusion?
	<ul style="list-style-type: none"> • Were patients, clinicians and study personnel kept "blind" to treatment received?
	<ul style="list-style-type: none"> • Were the groups similar at the start of the trial?
	<ul style="list-style-type: none"> • Was the trial stopped early?
Diagnosis	<ul style="list-style-type: none"> • Was there an independent, blind comparison with a reference standard?
	<ul style="list-style-type: none"> • Did the patient sample include an appropriate spectrum of the sort of patients to whom the diagnostic test will be applied in clinical practice?
	<ul style="list-style-type: none"> • Did the results of the test being evaluated influence the decision to perform the reference standard?
	<ul style="list-style-type: none"> • Were the tests methods described clearly enough to permit replication?
Harm	<ul style="list-style-type: none"> • Were there clearly identified comparison groups that were similar with respect to important determinant of outcome, other than the one of interest?
	<ul style="list-style-type: none"> • Were outcomes and exposures measured in the same way in the groups being compared?
	<ul style="list-style-type: none"> • Was follow-up of patients sufficiently long and complete?
	<ul style="list-style-type: none"> • Is the temporal relationship correct?
	<ul style="list-style-type: none"> • Is there a dose-response gradient?
Prognosis	<ul style="list-style-type: none"> • Was there a representative and well defined sample of patients at a similar point in the course of disease?
	<ul style="list-style-type: none"> • Was follow-up sufficiently long and complete?
	<ul style="list-style-type: none"> • Were objective and unbiased outcomes criteria used?
	<ul style="list-style-type: none"> • Was there an adjustment for important prognostic factors?
Systematic Review	<ul style="list-style-type: none"> • Did this review address a focused clinical question?
	<ul style="list-style-type: none"> • Were the criteria for article inclusion appropriate? (taking into account the type of question being asked)
	<ul style="list-style-type: none"> • Was the search for relevant studies exhaustive?
	<ul style="list-style-type: none"> • Was the validity of the included studies appraised?
	<ul style="list-style-type: none"> • Were the assessments of studies reproducible?
	<ul style="list-style-type: none"> • Were the results similar from study to study?
Practice Guidelines	<ul style="list-style-type: none"> • Were all important options and outcomes clearly specified?
	<ul style="list-style-type: none"> • Was an explicit and sensible process used to identify, select and combine evidence?
	<ul style="list-style-type: none"> • Was an explicit and sensible process used to consider the relative value of different outcomes?
	<ul style="list-style-type: none"> • Were the important recent developments included?
	<ul style="list-style-type: none"> • Has the guideline had peer review and testing?

What are the results?

Can you apply the results to your individual clinical question?

Table extracted from *User's Guide to the Medical Literature, Evidence-Based Medicine Working Group*

For All Types of Questions		<ul style="list-style-type: none"> • What are the overall results and the precision of the estimates? • Are the results applicable to your own individual population or patient? (Were the study patients similar to my own? Was the setting of the study applicable to my practice?)
Type of Question		<ul style="list-style-type: none"> • Considerations Specific to Particular Types of Questions
Therapy or Prevention	<i>results</i>	<ul style="list-style-type: none"> • To estimate the size of the Treatment effect, you want to look at Relative Risk, Odds Ratios or Numbers Needed to Treat to prevent adverse outcomes (See Survival Statistics Cheat Sheet)
	<i>applicability</i>	<ul style="list-style-type: none"> • Were all clinically relevant outcomes considered? • Are the benefits worth the harms and costs?
Diagnosis	<i>results</i>	<ul style="list-style-type: none"> • To estimate the ability of a test to change your pretest probability of disease, you want to look at Likelihood ratios (See Survival Statistics Cheat Sheet)
	<i>applicability</i>	<ul style="list-style-type: none"> • Will the test be reproducible and well interpreted in my practice setting? • Will the test results change my management? • Will my patients be better off because of the test?
Harm	<i>results</i>	<ul style="list-style-type: none"> • To estimate the strength of the association between the exposure and the outcome, you want to look at Relative Risk, Odds Ratios or Numbers Needed to Cause adverse outcomes (See Survival Statistics Cheat Sheet)
	<i>applicability</i>	<ul style="list-style-type: none"> • What is the magnitude of the risk? • Should I attempt to stop the exposure?
Prognosis	<i>results</i>	<ul style="list-style-type: none"> • To estimate the prognostic risk, you want to look at absolute risk (e.g. 5 yr. survival rate), relative risk (e.g. risk from a prognostic factor) or cumulative events over time (e.g. survival curves). • What are the possible outcomes and how likely are they to occur over time?
	<i>applicability</i>	<ul style="list-style-type: none"> • Will the results lead directly to selecting therapy? • Are the results useful for counseling patients?
Systematic Review	<i>results</i>	<ul style="list-style-type: none"> • What are the overall results when considering all of the studies reviewed and what is the precision of these results?
	<i>applicability</i>	<ul style="list-style-type: none"> • Specific Questions to determine whether you can apply these results to your population or patient should be determined by the type of question you are asking (e.g. Therapy vs. Diagnostic Testing, vs. Prognosis)
Practice Guidelines	<i>results</i>	<ul style="list-style-type: none"> • Are practical, important recommendations made? • How strong are the recommendations? • Could the uncertainty in the evidence or values change the guideline's recommendations
	<i>applicability</i>	<ul style="list-style-type: none"> • Is the objective of the guideline consistent with mine? • Are the recommendations applicable to my patients?

ⁱ Green ML. Evidence-based medicine training in internal medicine residency training programs. *J Gen Intern Med* 2000; 15: 129-133.

ⁱⁱ Richardson WS, Wilson MC, Nishikawa J, Hayward RSA. The well-built clinical question: a key to evidence based decisions. *ACP Journal Club*. Nov-Dec 1995; 123: A-12.

Article with an Example of Applying Evidence

Muir, AJ, Keitz, SA, and Schardt, CM. Applying evidence to the care of a patient with hepatitis C: How to develop a focused clinical question. [Seminars in Medical Practice Sept 2002 5\(3\):6-16.](#)

Clinical Question

The Clinical Question and Information Resources

Answering the Clinical Question: Critical Appraisal- Survival Skills

Define the Clinical Question.
1. Patient, Population or Problem
2. Intervention, Prognostic Factor, Exposure
3. Comparison (if appropriate)
4. Outcome you would like to measure or achieve
5. Type of Question you are asking
6. Type of Study you would want to find

As a fundamental part of your thinking about the elements of the clinical question, you need to decide what ‘type’ of question you are asking, as well as what kind of study you would love to find. This is because you will need to consider those questions when you are moving on to the next step of selecting and finding your resources.

What types of questions may we come up with?

Question Type	Possible Study Designs
1. Clinical Examination	Prospective cohort, blind comparison to Reference Standard
2. Diagnostic Testing	Prospective cohort, blind comparison to Reference Standard
3. Prognosis Series	Cohort Study (can be in the context of an RCT) > Case Control > Case
4. Therapy	RCT is really the only way we want to answer this question
5. Etiology / Harm	RCT (if possible and ethical) > Cohort Study > Case Control > Case Series
6. Prevention	RCT > Cohort Study > Case Control > Case Series
7. Cost	Economic Analysis
8. Self-Improvement/Education	
9. Quality Improvement	
10. Health Services Research	
11. Differential Diagnosis	

Question to Consider: Was the type of study the strongest that could have been performed under the circumstances?

Remember that it may not be either practical or ethical to use certain methodologies depending on the question. For example, it would not be ethical to randomize someone to a harmful treatment. Likewise, it may not be possible to do a prospective trial for an outcome that either takes years to develop or is very rare.

Types of Studies:

Experimental Design:
Randomized Control Trial (RCT) Randomization should ensure that comparison groups are equal. This is an experimental method.
Non- Experimental Design:
Cohort Study: follow one or more groups of individuals who have not yet suffered the adverse event and monitor the number of outcomes that occur over time. These need to be done when it is either not ethical or not practical to randomly assign patients to be “exposed” to something. Observational Design can be prospective or retrospective.
Case-Control Study: Collection of “cases” who have suffered the outcome and “controls” who have not. Investigators count the number of patients with a prognostic factor in the cases and the controls. These need to be done when the outcome of interest is rare or takes a long time to develop.
Case Series and Case Reports: Reports of patient scenarios that do not provide any comparison group.

PICO Table Handout

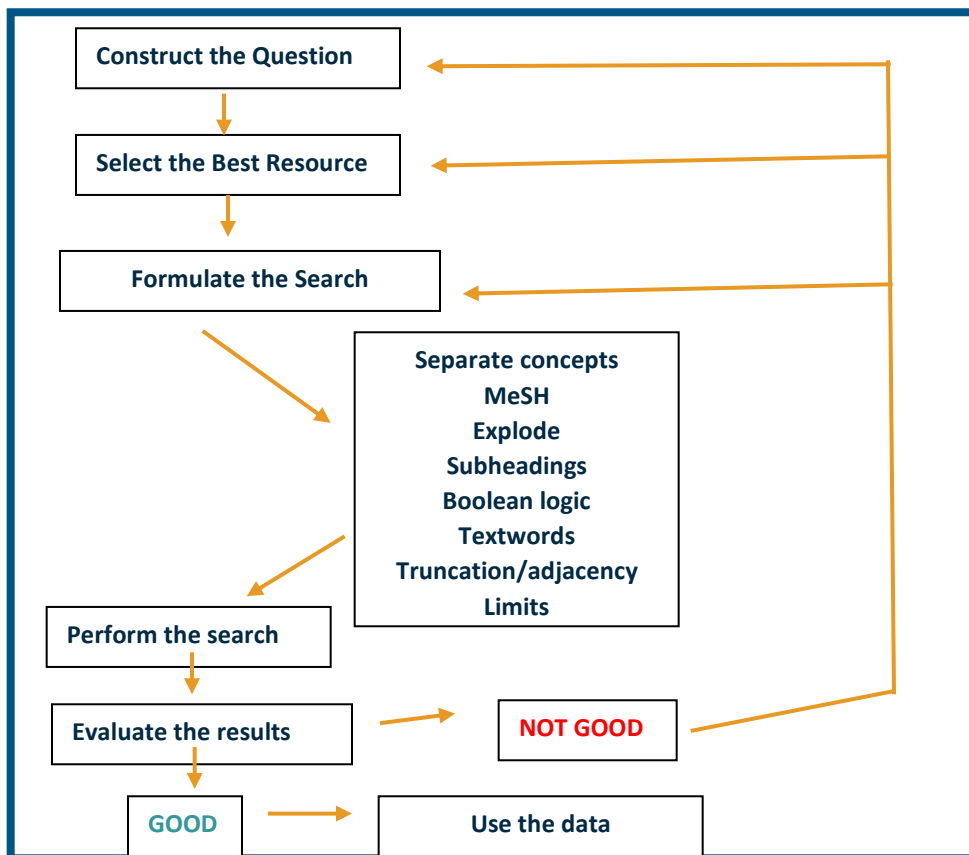
Patient, Population or Problem	How would I describe a group of patient's similar to mine	
Intervention, Prognostic Factor, Exposure	Which main intervention, exposure, or prognostic factor am I considering?	
Comparison (if appropriate)	What is the main alternative to compare?	
Outcome you would like to measure or achieve	What can I hope to accomplish, measure, improve, affect?	
Type of Question you are asking	How would I categorize this question?	
Type of Study you would want to find	What would be the best study design in order to answer this question?	

Patient, Population or Problem	How would I describe a group of patient's similar to mine	
Intervention, Prognostic Factor Exposure	Which main intervention, exposure, or prognostic factor am I considering?	
Comparison (if appropriate)	What is the main alternative to compare?	
Outcome you would like to measure or achieve	What can I hope to accomplish, measure, improve, affect?	
Type of Question you are asking	How would I categorize this question?	
Type of Study you would want to find	What would be the best study design in order to answer this question?	

Patient, Population or Problem	How would I describe a group of patient's similar to mine	
Intervention, Prognostic Factor Exposure	Which main intervention, exposure, or prognostic factor am I considering?	
Comparison (if appropriate)	What is the main alternative to compare?	
Outcome you would like to measure or achieve	What can I hope to accomplish, measure, improve, affect?	
Type of Question you are asking	How would I categorize this question?	
Type of Study you would want to find	What would be the best study design in order to answer this question?	

Searching for Evidence

Formulating Search Strategy



Parts of the Question:	Your question:	Your search strategy:
Patient Problem		
Intervention		
Comparison		
Outcome		
Type of Question		
Type of Study		

EBM Resources Comparison Chart

EBM Database/Resource	Access options	Content summary	Value/significance	More info/Bottom Line
ACP Journal Club American College of Physicians American Society of Internal Medicine (ACP-ASIM) acpj.acponline.org/index.html	<ul style="list-style-type: none"> - Abstracts included in PubMed (free) - Web access at acjc.acponline.org/ (\$) - In print or online in <i>Annals of Internal Medicine</i> (\$) - Via the OVID EBMR collection (\$) 	Includes value-added abstracts and commentary on selected original studies and systematic reviews. Focus is internal medicine.	Filtering of 100+ top journals for clinically relevant, methodologically sound studies. Expert commentaries on clinical usefulness supplement enhanced abstracts.	Excellent source to find important studies in medicine that are pre-appraised. JAMA. 2000; 283(14): 1875-1879
ACCESSSS From McMaster University https://www.accessss.org/	<ul style="list-style-type: none"> -Free to search -Registration required -Articles that are not open access must either be purchased or accessed through one's institution (option to link to your institution) 	Pre-appraised evidence to address this key question: what is the current best available evidence to support clinical decisions.	Sorts results based on the EBHC Evidence Pyramid 5.0 (systems, summary clinical texts, guidelines, systematic reviews, studies).	Sorting by evidence level and study design is helpful.
Cochrane Library www.cochrane.org	<ul style="list-style-type: none"> - In the Cochrane Library (\$) - Abstracts only at www.cochrane.org/ (free) - Abstracts in PubMed - Via the OVID EBMR collection (\$) 	Regularly updated collection of EBM databases, including: <ul style="list-style-type: none"> - Cochrane Database of Systematic Reviews - Cochrane Central Register of Controlled Trials - CDSR protocols 	The Cochrane Library aims to “prepare, maintain & promote the accessibility of systematic reviews of the effects of healthcare interventions.”	Within the Cochrane Library, it’s the Cochrane Database of Systematic Reviews that is of primary interest to clinicians. It sets the gold standard for quality reviews on clinically-relevant topics. J Med Libr Assoc. 2005 July; 93(3): 409–410
Cochrane Central Register of Controlled Trials (CENTRAL)	<ul style="list-style-type: none"> - In the Cochrane Library (\$) - In the OVID EBMR collection (\$) 	An international collection of controlled trials from a variety of sources.	Includes reports published in other sources not currently listed in MEDLINE or related databases.	Use the Cochrane Trials register when preparing a new systematic review or searching for clinical trials from the international literature.

Cochrane Database of Systematic Reviews	<ul style="list-style-type: none"> - In the Cochrane Library (\$) - Abstracts only at www.cochrane.org/ - Abstracts in PubMed - In the OVID EBMR collection (\$) 	Systematic reviews, most using meta-analysis, from the 50 Collaborative Review Groups. Focused topic summaries.	The gold standard for systematic reviews.	Use the Cochrane Database of Systematic Reviews to locate high quality, well-documented Systematic reviews.
DynaMed EBSCO www.dynamed.com	<ul style="list-style-type: none"> - Online (\$) - App available (\$) 	Focuses on diagnosis and treatment options. Systematic literature surveillance process monitors over 500 journals and integrates new evidence as it is published.	Extensive; updated daily. The bulleted format provides highlights from primary literature; users are encouraged to access original studies. Adheres to EBM principles – doesn't answer questions for you.	Use for finding the evidence at the point of care. Useful for tracking back to the original studies. J Med Libr Assoc. 2005 July; 93(3): 412–414
Essential Evidence Wiley www.essentialevidenceplus.com	<ul style="list-style-type: none"> - Online (\$) 	Filtered, synopsised, evidence-based information, allowing you to search across multiple databases, including EBM Guidelines, Daily POEMs, Cochrane Abstracts, Practice Guidelines, Calculators, Diagnostic Tests, and Calculators	Integrated system for quick searching of key resources; primarily of interest to family medicine practitioners.	Use Essential Evidence for “just-in-time” evidence-based answers to primary care clinical questions. Med Ref Serv Q. 2009 Spring; 28(1):105-6. DOI: 10.1080/02763860802616144
Evidence-Based Medicine Reviews (OVID EBMR) OVID www.ovid.com	<ul style="list-style-type: none"> - Online (\$) 	Single search engine for Cochrane Library files plus the ACP Journal Club.	Integrated searching of core EBM databases using the OVID interface.	Use OVID EBMR for a single portal to the main Cochrane Library files and ACP Journal Club
Google Scholar www.scholar.google.com	<ul style="list-style-type: none"> - Online (free) 	Searches broadly across a variety of web-based resources	Easy to use; familiar tool; does not disclose exactly what sites and resources it searches. Lacks good EBM filters.	When you can't find information in other databases N Engl J Med. 2006 Jan 5;354(1):4-7 J Med Libr Assoc. 2006 January; 94(1): 97–99

PubMed Clinical Queries National Library of Medicine https://www.ncbi.nlm.nih.gov/pubmed/clinical/	- Online (free) via PubMed – www.pubmed.gov	Filters for retrieving methodologically sound studies in four categories (therapy, diagnosis, etiology, and prognosis) plus systematic reviews.	Quick access for retrieving evidence-based original studies and systematic reviews from MEDLINE (based on the work of RB Haynes from McMaster).	Use PubMed’s clinical queries to select evidence-based studies from the MEDLINE database.
TRIP Database Centre for Research Support www.tripdatabase.com	- Online (free)	Meta-search engine for sources of high-quality internet information, including PubMed’s clinical queries, government guidelines, e-journals, and e-textbooks.	EBM-specific features such as a PICO search, evidence filters, and rapid review make this a relevant and interesting way to search for the best evidence.	Use TRIP when seeking pre-appraised evidence, reviews, and guidelines. Strong UK/Canada/Australia focus.
UpToDate www.uptodate.com	- Online (\$)	Concise, peer-reviewed topical summaries, chiefly in internal medicine and its subspecialties, focusing on diagnosis and treatment.	An easy-to-use database that provides quick answers to clinical questions. Summaries are a combination of synthesized literature reviews and expert knowledge.	Use UpToDate for peer-reviewed answers to specific clinical questions J Med Libr Assoc. 2003 January; 91(1): 97

PubMed Basics

PubMed[®] is the U.S. National Library of Medicine's (NLM) premier search system for health information. It is available **free** on the Internet at <http://mclibrary.duke.edu/pubmed>

PubMed Content

Over 33 million citations including:

- **Publisher-supplied citations** that will be analyzed to receive full indexing for MEDLINE if they are biomedical in nature
- **In-process citations** that have not yet been analyzed and indexed for MEDLINE[®]
- **Indexed for MEDLINE** citations of articles from about 5600 regularly indexed journals; MEDLINE makes up nearly 90% of PubMed.

PubMed Features

- Sophisticated search capabilities, including spell checker, Advanced Search Builder, and special tools for searching for clinical topics
- Assistance in finding search terms using the MeSH (Medical Subject Heading) database of MEDLINE's controlled vocabulary
- Ability to store citation collections and to receive email updates from saved searches using PubMed's My NCBI
- Links to full-text articles, to information about library holdings, and to other NLM databases and search interfaces



PubMed Searching

To search PubMed, type a word or phrase into the query box, including subject, author and/or journal. Then click on the **Search** button or press the Enter key. Combine search terms with connector words: “AND”, “OR” or “NOT” using upper case letters.

PubMed offers alternative searching options; for example, the Auto Suggest drop down menu appears when entering words and often a **Titles with your search terms** box is available after a search.

Limit searches by using the **Filters** list in the left navigation bar. Click on a term to activate or deactivate the filter. Use **Choose additional filters** for the full list of filters including Text availability, Publication dates, Species (Humans or Animals), Article types, Languages, Subjects, Ages, Sex, and Journal categories. Multiple choices may be made within sections. Make selections then click the **Search** button.

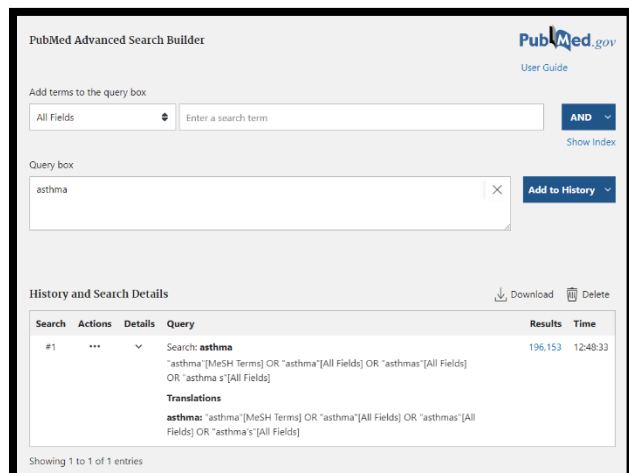
The **Filters activated** message appears above the search results list and limits remain in effect until removed or cleared.

Advanced Search

The Advanced search link provides two options to refine and focus a search: a Search Builder and Search History. The Advanced search box always defaults to Clear and does not retain previously run searches.

PubMed Advanced Search Builder offers the creation of a search using Boolean operators. Using the All Fields selection will run search terms through the Automatic Term Mapping process. A specific field may be selected from the drop-down menu to apply to the term. **Show index list** is available to display the search field index and the number of citations for each term in the search field. The Index display allows selection of multiple terms to “OR” together.

History and Search Details tracks search statements and numbers them, and also provides information on how PubMed interpreted the search (**Search Details, under >**).

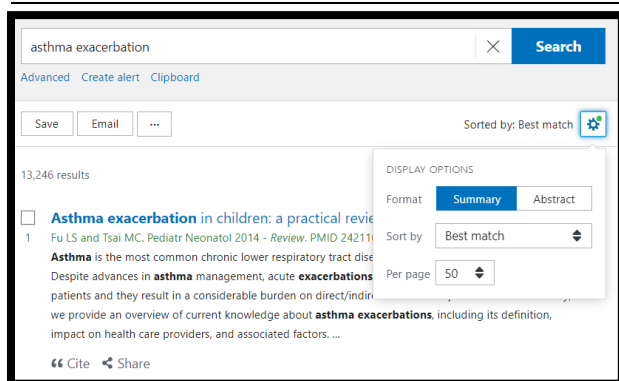


As mentioned, **Search details** provides information on how PubMed ran a search. PubMed looks first for the word or phrase as a MeSH term, then for journal titles, then authors. PubMed also searches “All Fields” for the word(s). **Search details** show how PubMed maps terms to MeSH headings (a process called Automatic Term Mapping).

Clipboard

This feature allows collection of selected citations from one or more searches for saving, printing, e-mailing, ordering, or storing in My NCBI Collections. The link is visible only when items are stored on the Clipboard. Click the check box next to citations to select them. Then click on the three ellipses next to Save and Email; select Clipboard from the dropdown list. To see the selected citations, click on the **Clipboard** link under the search box.

PubMed Search Results



After clicking on the **Search** button, PubMed displays a list of results in Summary format. To retrieve more information about the search results, use the **Sorted By** menu to change the view to the **Abstract** format.

PubMed defaults to showing the search results by **Best Match**, a relevancy ranking option. You can change the sort order to **Most Recent** or **Publication Date** here.

Similar Articles

A helpful PubMed feature is the ability to find citations that are similar to those of interest. To review Related Articles, click on a citation and scroll or click on the **Related Articles** link in the right hand menu on the Abstract view.

Sensors

PubMed examines search terms for certain elements and provides a shaded area above the search results with links to one or more citations or databases.

- **Citation Sensor:** matches search terms with citation elements (e.g. blood choi 2009)
- **Gene Sensor:** checks for the symbol of a gene found in the Gene database (e.g. CFTR)
- **Sequence Sensor:** detects accession numbers of nucleotides or proteins (e.g. X62176)
- **Structure Sensor:** detects items (proteins, etc.) in the Structure database (e.g. 1R10)

MeSH Database

Articles are indexed using a powerful vocabulary, called Medical Subject Headings (MeSH). The MeSH Database provides the option of identifying appropriate MeSH terms for searches.

The **MeSH Database** is available from the PubMed homepage. Use the MeSH database to search for a particular term or concept. If multiple items are retrieved, click on the desired term to view and select subheadings and other options. Then click on the **Add to Search Builder** button on the right side of the screen. When finished adding terms, click **Search PubMed** to complete the search.

The screenshot shows the 'PubMed Clinical Queries' search page. At the top, there is a search bar with the placeholder text 'Search PubMed' and a blue 'Search' button. Below the search bar, there are two columns of filters. The left column is titled 'COVID-19 Articles' and has a 'Category:' dropdown menu set to 'Treatment'. The right column is titled 'Clinical Study Categories' and has a 'Category:' dropdown menu set to 'Therapy' and a 'Scope:' dropdown menu set to 'Broad'. At the bottom of each column, there is a small note: 'This column displays citations about the 2019 novel coronavirus filtered by research topic categories. See more filter information.' and 'This column displays citations filtered to a specific clinical study category and scope. These search filters were developed by Haynes RB et al. See more filter information.'

Clinical Queries

PubMed's **Clinical Queries** section, accessed from the homepage or the Advanced Search page, makes it easier to find articles that report applied clinical research.

To search by **clinical study category**, enter search terms in the box provided. Then select a category: etiology, diagnosis, therapy, prognosis, or clinical

prediction guides and then choose either "broad" or "narrow" scope.

You can also access the Clinical Queries on the Duke PubMed results page, where the filters appear on the right of search results.

Saving, Emailing, and Downloading Results

After selecting your citations (i.e. from checked boxes or Clipboard), select **Save** or **Email**. The Save option allows you to send the citations via .RIS to citation management tools; you can also download a list of the PMIDs, or save the citations in a .CSV file.

Saving Searches in PubMed: My NCBI

Sign in or Register

Connect to <http://www.mclibrary.duke.edu/pubmed>. In the upper right corner of the screen click on **Log in**. If you have an account, you can sign in using your username and password. New users may register by clicking on **Register for an account**.

Save a Search | Get Updates via Email

My NCBI allows you to save search strategies. It can also deliver updates of search results to your email on a schedule that you determine.

Creating a Search Strategy

There are two ways to create a search strategy in PubMed.

1. Enter all terms into the search box, e.g. osteoarthritis AND (exercise OR exercise therapy).
2. On the **Advanced Search** page, build a strategy from the *History* using search numbers, e.g., #1, or click the *Add to builder* link corresponding with the relevant search sets.

#4 Search **#1 AND (#2 OR #3)**

#3 Search **exercise therapy**

#2 Search **exercise**

#1 Search **osteoarthritis**

Saving a Search Strategy

1. From the Results screen, click on **Create alert** below the search box. **NOTE:** The entire session history will not be saved, only the search statement that you are currently viewing. This search statement should include all relevant concepts.
2. Sign into **My NCBI**, if you are not already.
3. Review the search strategy for accuracy. Note that set numbers have been replaced by the terms searched.
4. Enter a new name for the search and click **Save**. **SUGGESTION:** Choose a name that is short and meaningful.
5. Select **No** or **Yes** to receive email updates.
6. If **Yes**, fill in the form indicating how often to get updates, the result format, and the number of items to send.
7. To access, delete, or edit settings of a search, sign into **My NCBI** and click on **Manage Saved Searches**.

Save Citations into My NCBI Collections

My NCBI also allows you to save individual citations indefinitely. You may create multiple collections within My NCBI for specific research projects that may be viewed privately or publicly shared with others.

Saving Selected Citations

1. After running a search, select the citations that you would like to save from the Results list by placing a checkmark in the box next to the citation.
2. Using the three ellipses next to Save and Email, select **Collections** from the dropdown. **Note: Don't choose My Bibliography**, as this is a separate function tied to the NIH open access policy, not a way to build a bibliography for a research topic.
3. Select whether you would like to create a new collection for the citations, which is the default option, or append (add) them to an existing collection.
4. Enter a name for your new collection or choose an existing collection from the drop-down menu. Click **Save**. **SUGGESTION:** If you choose to create a new collection, choose a name that is short and meaningful.
5. To access, delete, or share collections, sign into **My NCBI** and click on **Manage Collections**.

Setting Up Filters in PubMed: My NCBI

Sign in or Register

Connect to <http://www.mcclibrary.duke.edu/pubmed>. In the upper right corner of the screen, click on **Log in**. If you have a My NCBI account, you can sign in using your username and password. New users may register by clicking on **Register for an account**.

Filters

Filters allow you to group your search results by specific criteria, such as publication type or age groups. They appear at the top of the column to the right of your results. Clicking on a filter will display results limited by the selected filter.

Duke defaults: The Duke version of PubMed features: *Duke Medical Center Library* (most but not all of Duke holdings), *Core Clinical Journals*, *Nursing Journals*, *Review articles*, *Systematic Reviews*, *ACP Journal Club* and filters for the clinical queries, which limit to the best study type depending on the type of question being asked: *Therapy/RCT*, *Diagnosis*, *Prognosis*, *Cohort Studies* and *Etiology/Harm*.

Note: When you sign in, your preferences will override these defaults.

To Add a Filter from the Results Page

1. Click on **Manage Filters** at the bottom of the **Filter your results** box to the right of your results and sign in.
2. You can select from the lists of **Popular**, **LinkOut**, **Properties**, **Links**, or **Search** to find filters.
3. Check the filter you would like to add.
4. To go back to your results, click on **PubMed** under **Popular** at the bottom of the page, then choose **Advanced Search** to access your search history.

Examples of Useful Filters:

These filters can make your PubMed searches faster and more efficient. To apply a filter, simply search for the filter name and then place a checkmark next to it.

- **Duke Medical Center Library** limits searches to articles that the Duke Medical Library has access to in electronic or print format. We have access to additional articles than what appear with this filter, so be sure to check the "All" list in your results and follow [Get it @ Duke](#) links to see if we have a particular citation.
- **Clinical Queries** limits searches using the Clinical Query filter for selected question types, such as **Therapy/Narrow**. Learn more about clinical queries at http://www.ncbi.nlm.nih.gov/books/NBK3827/#pubmedhelp.Clinical_Queries_Filters.
- **NOTE:** Clinical queries are listed under **Properties**. Therapy/Narrow is recommended, as it filters your search results to look for randomized controlled trials. Other queries are also available.
- **Age Groups** limits searches to articles indexed to specific age groups, such as children or the aged.
- **Nursing Journals** limits searches to articles in nursing journals.

Create Your Own Custom Filters:

Want to filter results to a group of journals or a specific search strategy? Custom filters allow you to create and save filters based on search strategies, such as a set of journals or a group of drugs.

- **Tips:** First run a search that captures your custom filter strategy in PubMed, then copy the strategy from the Search. Details box on your search results and paste it into the custom filter area in My NCBI.

Therapy

Therapy Core Concepts

Notes on learning and teaching about **Therapy**:

1. **Background:** Many teachers and users of the medical literature will spend most of their time reading and teaching about randomized trials. Randomized methodology is intended to set up groups with equal prognosis that allow interventions to be compared on a level playing field. Many different types of questions are addressed using this study design including therapeutic interventions, screening, disease management strategies (e.g. intensive versus standard blood pressure control), or systems interventions (e.g. strategies to decrease readmissions).
2. Key concepts and terms
 - Concept of equal prognosis between groups
 - Randomization including intent of randomization and generation of random sequence (stratification and blocking)
 - Allocation Concealment
 - Blinding (masking)
 - Measures of treatment effect
 - Risk Ratio (also called Relative Risk)
 - Relative risk reduction or relative benefit increase
 - Risk Difference (also called Absolute Risk Reduction or Absolute Benefit Increase)
 - Number Needed to Treat (NNT)
 - Understanding precision: confidence intervals
 - Patients analyzed in the groups to which they were randomized (Intention to treat)
 - Follow-up
 - Applicability and generalizing results
 - Surrogate versus patient-important outcomes
 - Risk versus benefit considerations
 - Values and preferences in decision-making
3. Additional topics pertaining to therapy
 - Trials stopped early for benefit
 - Composite end points
 - Non-inferiority trials
 - Non-inferiority margin (incorporation of risk / benefit considerations)
 - Per-protocol analysis vs. Intention to treat analysis
 - Beware of faulty comparators (e.g. wrong dose or monitoring of the standard treatment) that make standard treatment look less effective

Therapy Critical Appraisal Form

Citation:

How serious is the risk of bias?	
Did intervention and control groups start with the same prognosis?	
Were patients randomized?	
Was randomization concealed?	
Were patients in the study groups similar at baseline with respect to prognostic factors?	
Was prognostic balance maintained as the study progressed?	
To what extent was the study blinded?	
Were groups prognostically balanced at the study's conclusion?	
Was follow-up complete?	
Were patients analyzed in the groups to which they were randomized?	
Was the trial stopped early?	
What are the results?	
How large was the treatment effect?	
How precise was the treatment effect?	
How can I apply the results to my patient care?	
Were the study patients similar to my patient?	
Were all patient-important outcomes considered?	
Are the likely benefits worth the potential harms and costs?	

Adapted from McMaster Evidence-based Clinical Practice Workshops and the Users' Guide to the Medical Literature 3rd Ed.

Therapy Critical Appraisal Form, Non-inferiority trials

Citation:

How serious is the risk of bias?	
Did intervention and control groups begin the study with a similar prognosis?	
Were patients randomized?	
Was randomization concealed?	
Were patients similar at baseline with respect to known prognostic factors?	
Was prognostic balance maintained as the study progressed?	
Were patients, caregivers, collectors of outcome data, adjudicators of outcome, and data analysts aware of group allocation?	
Were groups prognostically balanced at the study's conclusion?	
Was follow-up complete?	
Was the trial stopped early for benefit?	
Were patients analyzed in the groups to which they were randomized?	
Did the investigators guard against an unwarranted conclusion of non-inferiority?	
Was the effect of the standard treatment preserved?	
Did the investigators analyze patients according to the treatment they received, as well as to the groups to which they were assigned?	

What are the results?	
How large was the treatment effect?	
How precise was the estimate of the treatment effect?	
How can I apply the results to my patient care?	
Were the study patients similar to my patient?	
Were all patient-important outcomes considered?	
Are the likely advantages of the novel treatment worth the potential harms and costs?	

Adapted from McMaster Evidence-based Clinical Practice Workshops and the Users' Guide to the Medical Literature 3rd Ed.

FRISBE Therapy Critical Appraisal Worksheet with Key Learning Points

THERAPY STUDY	Article author/year: _____	Key Learning Points
<p>A. ARE THE RESULTS VALID? (“FRISBE”)</p> <p>F = Patient Follow-Up</p> <p>Were all patients who entered the trial properly accounted for and attributed at its conclusion? Was follow-up complete?</p>		<p><i>How do dropouts threaten validity?</i></p> <p>Study participants are lost to follow-up (LTF) when their status/outcomes are not known at the end of the trial. Often the reason that they are lost to follow-up relates to a systematic difference in their prognosis from those who continue with a study until the end (e.g. patients LTF do worse/are dead or may be greatly improved/ don’t feel the need to continue in the study). Thus, the loss of many participants may threaten validity.</p> <p>Furthermore, if the loss to follow-up is different between the two groups, dropouts or those lost to follow-up may create missing data that can disrupt the balance in groups created by randomization.</p>
<p>AR = Randomization</p> <p>Was the allocation (assignment) of patients to treatment randomized?</p> <p>Was the allocation concealed?</p>		<p><i>Why is randomization important?</i></p> <p>Effective randomization guarantees that each subject has an independent and fixed chance of being allocated to each group. The chance is usually equal (e.g. in parallel group design where a participant is randomized to one of two or more interventions).</p> <p>Randomization aims to balance groups for known and unknown prognostic factors by allocating subjects to groups by chance alone. If randomization is correctly done, any group differences should be attributable to chance alone. The intent is to minimize chance differences so that any observed group differences can be attributed to the effect of treatment.</p> <p>Allocation concealment assures that those assessing eligibility and assigning subjects groups don’t have knowledge of the allocation sequence.</p>

<p>I = Intention-to-Treat Analysis</p> <p>Were patients analyzed in the groups to which they were randomized?</p> <p>Were all randomized patient data analyzed?</p>		<p><i>Why is intention-to-treat analysis important?</i></p> <p>ITT preserves the balance of prognostic factors in groups created by the original random group allocation. It provides the truest estimate of the effects of treatment allocation in real-world practice by including data from crossovers, non-adherents, dropouts and those lost to follow-up, plus estimates of missing data points. ITT thereby avoids overly optimistic estimates of treatment efficacy resulting from excluding non-compliers.</p>
<p>S = Similar Baseline</p> <p>Characteristics of Patients Were groups similar at the start of the trial?</p>		<p><i>Why should groups be similar at baseline?</i></p> <p>It is important to verify that those factors known to influence outcome are equally distributed. And to assess the potential effect on the study outcome of an imbalance that occurs by chance.</p>
<p>B = Blinding</p> <p>Were patients, health workers, and study personnel "blind" to treatment?</p>	<p>Blinded groups included (Y=yes, N=no, U=uncertain):</p> <p>_____ patients</p> <p>_____ providers</p> <p>_____ raters or assessors</p> <p>_____ data analysts</p> <p>_____ adjudicators</p>	<p><i>Why is blinding important?</i></p> <p>Blinding equalizes the effect of patient and provider expectations on outcome across groups. For raters, blinding minimizes subjectivity in outcome measurement. For providers, blinding eliminates the possibility of either conscious/unconscious differential administration of effective intervention to either group: i.e. co-interventions (unintended additional care to either group) or contamination (provision of intervention to control group).</p>
<p>E = Equal Treatment</p> <p>Aside from the experimental intervention, were the groups treated equally?</p>		<p><i>Why should groups be treated equally?</i></p> <p>Equal treatment helps guarantee that the groups will remain prognostically balanced by avoiding systematic differences in the care provided other than the intervention.</p>
<p>Summary of article's validity</p>	<p>Notable strengths / weaknesses:</p> <p>Overall, this trial method is (strong/adequate/weak)</p> <p>Potential threats are (minimal/modest/serious/fatal) and would likely bias the results of the study towards (overestimate/underestimate) of treatment effect.</p>	<p>How serious are the threats to validity and in what <i>direction</i> could they bias the study outcomes?</p> <p>Include notable strengths/weaknesses as well as the direction of the biases and how that may impact interpretation of results.</p>

<p>B. WHAT ARE THE RESULTS?</p> <p>How large was the treatment effect?</p> <p>How precise was the treatment effect?</p>	<p>1) Response rates on dichotomous outcome measure:</p> <table border="1" data-bbox="619 235 1260 409"> <thead> <tr> <th data-bbox="619 235 747 293">Outcome</th> <th data-bbox="747 235 842 293">EER¹ (n=)</th> <th data-bbox="842 235 987 293">CER or EER² (n=)</th> <th data-bbox="987 235 1131 293">Risk Difference</th> <th data-bbox="1131 235 1260 293">NNT (95%CI)</th> </tr> </thead> <tbody> <tr> <td data-bbox="619 293 747 409"></td> <td data-bbox="747 293 842 409"></td> <td data-bbox="842 293 987 409"></td> <td data-bbox="987 293 1131 409"></td> <td data-bbox="1131 293 1260 409"></td> </tr> </tbody> </table> <p>Risk Ratio</p> <p>Risk Difference</p> <p>NNT or NNH</p>	Outcome	EER ¹ (n=)	CER or EER ² (n=)	Risk Difference	NNT (95%CI)						<p>Calculate and state the plain English meaning of summary statistics for dichotomous outcomes:</p> <p>Risk Ratio</p> <p>Risk Difference</p> <p>NNT or NNH</p>
Outcome	EER ¹ (n=)	CER or EER ² (n=)	Risk Difference	NNT (95%CI)								
<p>C. WILL THE RESULTS HELP ME IN CARING FOR MY PATIENTS?</p> <p>Can the results be applied to my patient?</p> <p>Were all clinically important outcomes considered?</p> <p>Are the likely treatment benefits worth the potential harms and costs?</p>												

Therapy Formulas

		Outcome		
		Outcome Present	Outcome Absent	
Treated/ Exposed (Y)	a Outcome present in treated patient	b Outcome absent in treated patient	Y= Risk of Outcome in Treated Group =a/(a+b)	
Control/ Not exposed (X)	c Outcome present in control patient	d Outcome absent in control patient	X= Risk of Outcome in Control Group =c/(c+d)	

I. Relative Risk

The ratio of risk of outcome in treated group (Y) as compared with control group (X)

$$RR=Y/X = a/(a+b) / c (c+ d)$$

This always tells us whether the observed outcome (effect) occurs more or less often in the exposed group than in the unexposed group. Calculations for RR are identical whether you are asking a question about therapy or a question about Harm. Relative Risk can only be calculated from RCTs or cohort studies where we can determine outcomes of interest in exposed / treated groups and unexposed / control groups. (Note: for case-control studies, the numbers of cases and controls and, therefore, the proportion of individuals with the outcome is chosen by the investigator- for case-control studies we use odds ratios: Odds Ratio = (a/c) / (b/d)

II. Relative Risk Reduction

The percent reduction is percent decrease in risk in the treated group (Y) as compared with control group (X)

$$RRR= [X - Y] / X \text{ or } 1- RR$$

For Questions of Harm: You calculate the Relative Risk Increase: The calculation is exactly the same as for Treatment, however, you will have an increase in relative risk.

III. Absolute Risk Reduction

The difference in risk between the control group (X) and the treated group (Y). The risk is higher in the control group, therefore, the you subtract

$$ARR = X (\text{Control}) - Y (\text{Treated})$$

For Questions of Harm: You calculate the Absolute Risk Increase. Because the risk is higher in the treated group, the ARI= Y(Treated)- X (Control)

IV. Number Needed to Treat

NNT is the reciprocal of the ARR **NNT = 1/ARR = 1/(X-Y)**
(an NNT of 20 means that 20 patients must be treated to prevent one adverse outcome)

For Questions of Harm: You calculate the Number Needed to Harm: The calculation is exactly the same as for Treatment, however, you will take the reciprocal of Absolute Risk Increase: **NNH=1/ARI = 1/(X-Y)**
(a NNH of 20 means that for every 20 patients treated, we will cause one adverse outcome)

Therapy Example Exercise

Therapy: To Dig or not to Dig- that is the question.

Overall goal: to define a strategy for finding and using an article that relates to therapy.

Article to be reviewed: The Digitalis Study Group. The effect of digoxin on mortality and morbidity in patients with heart failure. [NEJM 1997;336:525-33](#)

Skills to be acquired:

1. *Ask:* Create a pertinent, answerable question from a clinical case scenario
2. *Acquire:* Perform a search in order to identify original research that will address your question
3. *Appraise:* Determine whether the article gives us valid information
4. *Apply:* Determine whether the results are applicable to the patient in your case

Specific Vocabulary and Tasks for this lesson. In addition to completing the critical appraisal sheet for reading a therapy article, please consider the following:

1. Randomization (including issues of concealment and stratification)
2. Blinding vs. Open Label
3. Intention to treat (methods page 526)
4. Sensitivity Analysis (see paragraph in methods on page 527)
5. Kaplan-Meier Analysis (Figures 1 and 2)
6. Please describe the results in terms of Relative Risk Reduction, Absolute Risk Reduction and Number Needed to Treat (NNT).

Clinical Case Scenario:

You are a second-year resident picking up a ward service at the VA. On rounds your first morning, your medical student presents the following patient to you:

The patient is a 73-year-old white man who was admitted with an exacerbation of CHF. He has a history of past MI and an EF by Echo of 25%. He is very compliant and seldom, if ever, misses his medications. On admission, his medications were furosemide 40 mg po qd, enalapril 20 mg po bid, simvastatin 10 mg po qd and aspirin 325 mg po qd. He usually functions at a NYHA class II level, but he now has had two admissions for CHF in the past 4 months. On this admission, he has already been ruled out for MI and remains in sinus rhythm. He has a normal renal function with a creatinine of 0.7. The patient is particularly concerned about staying out of the hospital because he lives alone and has no one to care for his pets or plants. He is very active in his community and is eager to get out and stay out of the hospital.

The medical student asks whether Digoxin would help this man with his heart failure. The student was told by one of his prior residents that Digoxin was a good choice for a patient with CHF. Your intern states that she feels that Digoxin does not improve mortality in patients with heart failure and that we must consider potential toxicities, especially in a 73-year-old man. They look to you for guidance.

Therapy Example Exercise with ANSWERS

Citation: The Digitalis Study Group. The effect of digoxin on mortality and morbidity in patients with heart failure. [NEJM 1997;336:525-33](#)

Are the Results Valid?	
Did experimental and control groups begin the study with a similar prognosis?	
Were patients randomized?	Yes. The randomization was stratified by LV function as well as by study site (302 sites).
Was randomization concealed?	Yes. The randomization was done through a coordination center and relayed to sites via telephone (i.e. the enrolling folks did not have any information about the randomization scheme.)
Were patients analyzed in the groups to which they were randomized?	Yes. The analysis was an intention to treat.
Were patients in the treatment and control groups similar with respect to known prognostic factors?	Yes. Table 1 on page 526 shows that the groups were equal at the start of the trial with respect to important prognostic factors.
Did experimental and control groups retain a similar prognosis after the study started?	
Were 5 important groups (patients, caregivers, collectors of outcome data, adjudicators of outcome, data analysts) aware of group allocation?	<ol style="list-style-type: none"> 1. patients: yes, to some extent. The study began with placebo controls and patients blinded to study arm. However, in some patients (if clinical need arose) a switch was made to open-label treatment. Measures were taken to minimize unblinding for patients and caregivers. However, there was an acknowledgement that this was not always possible. 2. caregivers: yes, with the same caveat as above. 3. collectors of outcome data: yes- blinding was protected, even if clinicians were unblinded 4. adjudicators: yes- blinding was protected, even if clinicians were unblinded 5. data analysts: yes- blinding was protected, even if clinicians were unblinded
Aside from the experimental intervention, were groups treated equally?	Yes. There are no differences that we know of in important co-interventions. Specifically, large proportions of patients in both groups received an ACE-inhibitors and diuretics in addition to Dig (94% vs. 81%)
Was follow-up complete?	Yes. The final status of 98.6% of patients is known at the end of the trial. In addition, the investigators comment on a sensitivity analysis (how sensitive the results would be to the 'worst' case scenario if all of the lost patients had the worst possible outcome). In this case, a sensitivity analysis showed that the 'lost' patients would not affect the overall mortality results, even if all had died.

What are the results?

How large was the treatment effect?

Mortality was the primary outcome measure; however, the hospitalization outcomes were also important. Results are summarized in the table below:

Outcome	Digoxin	Placebo	RRR	ARR	NNT
Mortality	34.8%	35.1%	Nonsignificant		(p=.08%)
Total hospital	64.3%	67.1%	35%	2.8%	36
CHF hospital	27%	35%	23%	8%	13
CV hospital	49.9%	54.4%	8.3%	4.5%	22

Note: Of course, suspected digoxin toxicity was a more frequent cause of hospitalization in the treated group.

How precise was the treatment effect?

95% Confidence Intervals are reported in the tables surrounding the risk ratios. Because the sample size is large for this trial, the CI's are fairly tight in general.

Table 2 (page 528) shows CI's for the Risk Ratios. The major outcomes (all-cause, cardiovascular and heart failure) all include 1 (i.e. no difference)

For hospitalization outcomes in Table 3, however, (page 529) several important outcomes have CI's that do not include 1 and favor the treated group.

How can I apply the results to my patient care?

Were the study patients similar to my patient?

In some ways, yes: 86% of the patients in the trial were white, 84% were either NYHA class II or III, and 70% of patient had ischemic cardiomyopathy. However, only 27% were older than 70 years. All in all, however, it is likely that our patient is close enough in characteristics to the study population that the results may be applied to him.

In addition, there are some ways in which the study results are particularly helpful to our patient. Notably, in our patient the outcome of preventing hospitalizations may be of great interest to him. She explicitly reports that staying out of the hospital is an important goal for her (and her pets!)

Were all patient-important outcomes considered?

Yes. Most (if not all) clinically relevant outcomes were considered.

Are the likely benefits worth the potential harms and costs?

This may be a matter of weighing patient's values. For a patient who elects to be treated with digoxin, there will be the need to monitor therapy, draw frequent drug levels, and hold the risk of toxicity. However, for some, these issues will be offset by the possible benefit of avoidance of hospitalization. From a resource utilization point of view, the cost of the drug (cheap) and monitoring (not quite so cheap) will still be less than a single hospitalization.

It is fairly convincing from this large, well-designed trial that there is not a large difference in mortality overall when using digoxin. Thus, the quality of life issues (including hospitalization) should predominate the discussion.

Adapted from McMaster Evidence-based Clinical Practice Workshops

Harm

Harm Core Concepts

Notes on learning and teaching about **Harm**:

1. **Background:** Questions of Harm are answered by a diverse set of study designs. For some questions of harm (e.g., harms associated with a particular therapy), harm can be studied in the context of RCTs simply by measuring harms in addition to benefits in your trial (e.g. an RCT studying thrombolytics would measure improvement in mortality as well as an increase in bleeding). However, when outcomes are rare, or studying them would be unethical, other study designs are frequently required. The context of harm can allow learners and teachers of the medical literature to gain an appreciation of the hierarchy of evidence (understanding the strengths and potential biases associated with different study designs to provide a gradation of more to less bias).
2. Key concepts and terms
 - Hierarchy of evidence
 - Study design ranging from experimental (RCT) to observational (cohort, case-control, case series)
 - Direction of inquiry
 - Odds vs. Risk and when each can be used
 - Association versus causality

Harm Critical Appraisal Form, Cohort Study

Citation:

How serious is the risk of bias?	
Aside from the exposure of interest, did the exposed and control groups start and finish with the same risk for the outcome?	
Were the patients similar for prognostic factors that are known to be associated with the outcome (or did statistical adjustment level the playing field)?	
Were the circumstances and methods for detecting the outcome similar?	
Was the follow-up sufficiently complete?	
What are the Results?	
How strong is the association between exposure and outcome?	
How precise is the estimate of risk?	
How can I apply the results to my patient care?	
Were the study patients similar to patients in my practice?	
Was follow-up sufficiently long?	
Is the exposure similar to what might occur in my patient?	
What is the magnitude of the risk?	
Are there any benefits that are known to be associated with the exposure?	

Adapted from McMaster Evidence-based Clinical Practice Workshops and Users' Guide to the Medical Literature 3rd Ed.

Harm Critical Appraisal Form, Case-Control Study

Citation:

How serious is the risk of bias?	
Did the cases and control group have the same risk (chance) for being exposed in the past?	
Were cases and controls similar with respect to the indication or circumstances that would lead to exposure?	
Were the circumstances and methods for determining exposure similar for cases and controls?	
What are the Results?	
How strong is the association between exposure and outcome?	
How precise is the estimate of risk?	
How can I apply the results to my patient care?	
Were the study patients similar to patients in my practice?	
Was follow-up sufficiently long?	
Is the exposure similar to what might occur in my patient?	
What is the magnitude of the risk?	
Are there benefits that offset the risks of the exposure?	

Adapted from McMaster Evidence-based Clinical Practice Workshops and Users' Guide to the Medical Literature 3rd Ed.

Harm Sample Packet: The Class Study (RCT)

Learning Objectives:

1. To apply the results of papers which assess issues of harm to clinical practice. This includes:
 - Determining the validity of the results (whether the results are true)
 - Interpreting the results (understanding the magnitude and precision of the effect)
 - Applying the results to your patients and clinical practice
2. To consider the different kinds of studies that might apply to questions of HARM (case-control, cohort, and RCT)
3. To consider possible sources of bias in the reporting process including the role of financial sponsorship of clinical research.

Enclosed Materials

1. Silverstein, FE, Faich, G, Goldstein, JL et al. Gastrointestinal toxicity with Celecoxib vs. nonsteroidal anti-inflammatory drugs for Osteoarthritis and Rheumatoid Arthritis. The CLASS Study: a randomized controlled trial. [JAMA. 2000; 284:1247-1255.](#)
2. **Part I of Educational Package:**
 - Clinical Scenario
 - Teaching Scenario
 - Critical Appraisal Worksheets
3. **Part II of Educational Package:**
 - Follow up information

Note:
**Please fully complete Part I prior to
reading or working on
Part II of this package**

Educational Package: Part I

Clinical Scenario:

HPI:

This is the first visit for Mr. G, a 64-year-old man with hypertension, diabetes, and osteoarthritis of the right knee and shoulders. He also has a remote history of peptic ulcer disease but has not taken medication for PUD for several years. He has recently retired to this area and no longer has insurance to cover drug costs. He admits that he is having difficulty in affording his medications. On arrival to your clinic, he states that his medical regimen includes carvedilol (12.5 mg po bid; US \$90.67 for a one-month supply), pioglitazone (15 mg po qd; US \$82.95 for a one-month supply) and Celebrex (100 mg po bid; US \$82.09 for a one-month supply). His prior physician told him that these medications were the best he could buy and the patient reports that he has done well on this regimen. Mr. G reports compliance with medications up to this point, however, he is sure he will no longer be able to purchase these medications on his fixed income.

Physical Exam:

Temp 98.3 BP 130/68. The patient is well appearing and in no distress. Examination of the heart, lungs and abdomen are unremarkable. Diabetic foot exam reveals normal sensation and 2+ pulses bilaterally.

Data:

Prior to your visit, Mr. G went to the lab. Blood chemistries are within normal limits and HgbA1C today is 6.8. He has no protein in his urine.

At the close of the visit, you praise Mr. G for his compliance and dedication to his health. You acknowledge that his medical problems are extremely well controlled, however, you are worried about his ability to pay for these meds. You suggest that there are alternative medications that can manage his medical conditions for much less cost and suggest a transition to a regimen that he can afford. As a start, you suggest hydrochlorothiazide (25 mg po qd; US \$8.00 per month), glipizide XL (5 mg po qd; US \$12.00 per month) and ibuprofen (600 mg po tid; purchased over the counter: 200 mg tabs; 500 tablets costs US \$6.00).

Mr. G agrees to switch to the hydrochlorothiazide and glipizide and appreciates the cost savings. However, he was told that the Celebrex would lower his risk of an upper GI bleeding complication. He is reluctant to make this switch. He asks you if you think there is a difference in the toxicity profiles of ibuprofen as compared with Celebrex.

You tell Mr. G that you will review the medical literature on this topic prior to his follow-up visit. Mr. G would like to continue the Celebrex until your next discussion.

Clinical Question:

In patients with Osteoarthritis treated with Cyclooxygenase inhibitors, does Celecoxib (Celebrex) cause fewer GI complications than ibuprofen?

Teaching Scenario:

That afternoon, you are joined by a third-year medical student who is shadowing you in the clinic for one-half day each week. After seeing another patient with osteoarthritis, you tell him about Mr. G’s dilemma and walk him through the formulation of the clinical question regarding HARM. The student questions the feasibility of studying a question regarding HARM. He raises the following outstanding questions:

- How can you go about studying a question regarding HARM?
- What do you do when the outcomes are infrequent?
- Is it ethical to give patients medications when you expect the exposures to cause HARM?

You look at your watch and see that you are beginning to run behind schedule. Therefore, you suggest to the student that he consider the following assignment prior to his return next week. You ask him to learn about 3 different study designs that might be used to answer a question of HARM, a case-control study, a cohort study and a randomized controlled trial. For each study design, you ask him to consider the following characteristics.

Study Design	How you select your starting population of patients	Comparison group	Assessments and direction of inquiry	Strengths / Weaknesses
1) Case-Control				
2) Cohort				
3) RCT				

In the spirit of shared learning, you tell him that if he brings the information on the study designs next week, you will bring a relevant paper to review.

Finding the information:

Remember the Clinical Question:

In patients with Osteoarthritis treated with Cyclooxygenase inhibitors, does Celecoxib (Celebrex) cause fewer GI complications than ibuprofen?

Your first strategy is to look this up in UpToDate. Within minutes, you find an overview of selective Cox-2 inhibitors and scan for the answer to your question. You find the following comment in the subsection entitled 'Reduction in gastroduodenal toxicity':

Celecoxib was also associated with significantly fewer symptomatic ulcers and ulcer complications compared with ibuprofen and diclofenac.

The reference for this statement is listed as:

Gastrointestinal Toxicity with Celecoxib vs Nonsteroidal Anti-inflammatory Drugs for Osteoarthritis and Rheumatoid Arthritis. The CLASS Study: A Randomized Controlled Trial. [JAMA 2000; 284: 1247-1255.](#)

(Note: For those of you who might prefer to search this clinical question in PubMed, it can be easily done! The following search was done in less than 3 minutes and produced the same exact article! Note that lines 1→3 of the search are directly pulled from elements of the clinical question. Lines 4 combines them together. Line 5 includes a filter for randomized controlled trials.

The search:

#5	Add	Search ((osteoarthritis) AND Cyclooxygenase inhibitors) AND peptic ulcer Filters: Randomized Controlled Trial	53
#4	Add	Search ((osteoarthritis) AND Cyclooxygenase inhibitors) AND peptic ulcer	179
#3	Add	Search peptic ulcer	83905
#2	Add	Search Cyclooxygenase inhibitors	125417
#1	Add	Search osteoarthritis	65292

Assessment of Validity

Having identified an article that might give us some insight into this question, we now would like to critically appraise the article. Because the clinical question is one regarding HARM, the critical review form for HARM is included. Because the paper is a randomized control trial, some of the questions that pertain to the review of a therapy question are also relevant. Therefore, a critical review form for THERAPY is included as well.

Please complete the critical review process for the paper and decide what you want to tell Mr. G.

Critical Review Form for Harm

Citation: Silverstein FE. et. al. Gastrointestinal toxicity with celecoxib vs. nonsteroidal anti-inflammatory drugs for osteoarthritis and rheumatoid arthritis. [JAMA. 2000;284:1247-1255](https://doi.org/10.1093/ajph/2000.284.1247-1255)

Are the Results Valid?	
Did experimental and control groups begin the study with a similar prognosis?	
Did the investigators demonstrate similarity with respect to all known determinants of outcome; did they adjust for differences in the analysis?	
For case-control studies: Were exposed patients equally likely to be identified in the two groups?	
Did experimental and control groups retain a similar prognosis after the study started?	
Were the outcomes and exposures measured in the same way in the groups being compared?	
Was follow-up sufficiently complete?	
What are the Results? How strong is the association between exposure and outcome? How precise is the estimate of risk?	
How can I apply the results to my patient care?	
Were the study patients similar to patients in my practice?	
Was the duration of follow-up adequate?	
Should I attempt to stop the exposure?	

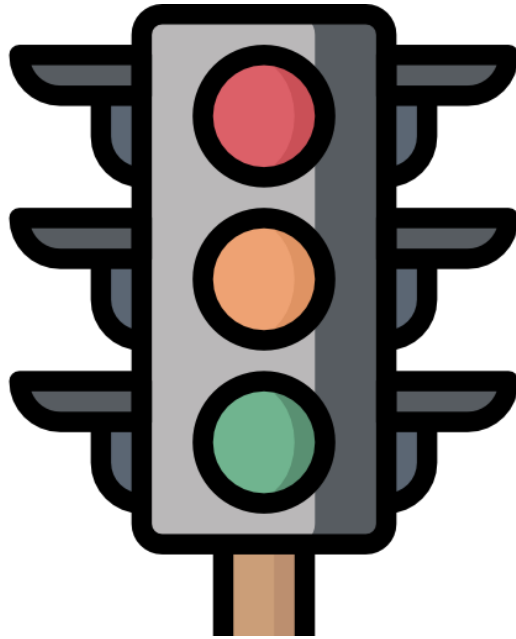
Adapted from McMaster Evidence-based Clinical Practice Workshops 2nd Ed.

Critical Review Form for Therapy

Citation: Silverstein FE. et. al. Gastrointestinal toxicity with celecoxib vs. nonsteroidal anti-inflammatory drugs for osteoarthritis and rheumatoid arthritis. [JAMA. 2000;284:1247-1255](#)

Are the Results Valid?	
Did experimental and control groups begin the study with a similar prognosis?	
Were patients randomized?	
Was randomization concealed?	
Were patients analyzed in the groups to which they were randomized?	
Were patients in the treatment and control groups similar with respect to known prognostic factors?	
Did experimental and control groups retain a similar prognosis after the study started?	
Were 5 important groups (patients, caregivers, collectors of outcome data, adjudicators of outcome, data analysts) aware of group allocation?	
Aside from the experimental intervention, were groups treated equally?	
Was follow-up complete?	
What are the Results? How large was the treatment effect? How precise was the treatment effect?	
How can I apply the results to my patient care?	
Were the study patients similar to my patient?	
Were all patient-important outcomes considered?	
Are the likely benefits worth the potential harms and costs?	

Adapted from McMaster Evidence-based Clinical Practice Workshops 2nd Ed.



**This concludes Part I of the Educational Package.
Please be sure you have completed the critical appraisal
exercise prior to moving on.**

**Specifically, decide whether you will advise Mr. G to pay the extra \$\$ for
celecoxib (pick one response below)**

Yes, he should continue the celecoxib

No, he should not continue the celecoxib

Once you have made this decision, you may move on to Part II.

Educational Package: Part II

The Rest of the Story...

The weekend after you see Mr. G, you happen to take a trip to Washington DC to participate in a National meeting of Lovers of EBM. You arrive on Sunday morning and are enjoying coffee and a bagel while reading the Washington Post. The following article catches your eye.

Missing Data on Celebrex

Full Study Altered Picture of Drug by Susan Okie

Washington Post Staff Writer

Sunday, August 5, 2001; Page A11

When editors of the Journal of the American Medical Association sent medical expert M. Michael Wolfe an unpublished study on the blockbuster arthritis drug Celebrex last summer, he was impressed by what he read.

Tested for six months in a company-sponsored study involving more than 8,000 patients, the drug was associated with lower rates of stomach and intestinal ulcers and their complications than two older arthritis medicines -- diclofenac and ibuprofen. JAMA's editors wanted to rush the findings into print, and Wolfe and a colleague provided a cautiously favorable editorial to accompany it. But in February, when Wolfe was shown the complete data from the same study as a member of the Food and Drug Administration's arthritis advisory committee, he said he saw a different picture.

"We were flabbergasted," he said.

The study -- already completed at the time he wrote the editorial -- had lasted a year, not six months as he had thought, Wolfe learned. Almost all of the ulcer complications that occurred during the second half of the study were in Celebrex users. When all of the data were considered, most of Celebrex's apparent safety advantage disappeared.

"I am furious. . . . I wrote the editorial. I looked like a fool," said Wolfe, a Boston University gastroenterologist. "But . . . all I had available to me was the data presented in the article."

JAMA's editor, Catherine D. DeAngelis, said the journal's editors were not informed about the missing data. "I am disheartened to hear that they had those data at the time that they submitted [the manuscript] to us," she said. "We are functioning on a level of trust that was, perhaps, broken."

The study's 16 authors included faculty members of eight medical schools. All authors were either employees of Pharmacia, Celebrex's manufacturer, or paid consultants of the company. For company-sponsored studies, JAMA now requires a statement, signed by an author who is not employed by the company, taking "responsibility for the integrity of the data and the accuracy of the data analyzes," DeAngelis added.

Steven Geis, a vice president for clinical research of Pharmacia and one of the authors, said that only the first six months of data were presented because, after that, more patients withdrew from the comparison groups than from the Celebrex group, biasing later findings. He said a three-member executive committee, composed of authors who were not Pharmacia employees, approved the decision.

"The intention really was not to be deceptive in any way," he said. "People thought that six months was the appropriate analysis." With the inclusion of the later data, "the actual difference between Celebrex and [the other drugs] are not as wide as they were at six months," he acknowledged. "But I think in the end, it does show that Celebrex has a superior safety profile."

After reviewing the full study, the FDA's arthritis advisory committee concluded that Celebrex offers no proven safety advantage over the two older drugs in reducing the risk of ulcer complications, said FDA spokesman Susan Cruzan. The company has requested a change in the drug's labeling to state that it is indeed safer, but the FDA has asked for additional information before making a decision.

Meanwhile, the JAMA article and editorial have likely contributed to Celebrex's huge sales. "When the JAMA article comes out and confirms the hype, that probably has more impact than our labeling does," said Robert J. Temple, director of medical policy at the FDA's Center for Drug Evaluation and Research.

James Wright, a professor of clinical pharmacology at the University of British Columbia, said he complained to JAMA after noticing differences between the published report and the data presented to the FDA. He praised the Public Citizen's Health Research Group, a consumer organization, for filing a lawsuit that led to the agency's putting all drug studies presented to its advisory committees on its public Web site.

"Otherwise, we still wouldn't know this," Wright said. "We would still be in the dark."

© 2001 The Washington Post Company

The FDA Data:

On return home, you are determined to view the full dataset. You eventually find the data at <http://www.fda.gov/ohrms/dockets/ac/01/briefing/3677b1.htm>

Some of this data is summarized in the tables below. Note that data are not available regarding person-years on the FDA website. Thus, from the web, we can only calculate risk over the 6-month or 12-month period. Statistics shown represent Log-Rank P-values. Also note that the NSAID data (n= 3981) can be broken down into Diclofenac (n=1996) + Ibuprofen (n=1985)

Six Month Data from the FDA Web site (risk, censored)

Twelve Month Data from the FDA Web site (risk, uncensored)

	# of events	Denominator (# of people)	Event rate	Relative Risk		# of events	Denominator (# of people)	Event rate	Relative Risk
6 mo. risk					12 mo. risk				
Celebrex	11	3987	0.28%		Celebrex	17	3987	0.43%	
NSAID	20	3981	0.50%	0.56 (p=0.092)	NSAID	21	3981	0.53%	0.81 (p=0.640)
NSAID Data from 6 months is broken into the 2 groups below:					NSAID Data from 12 months is broken into the 2 groups below:				
Diclofenac	9	1996	0.45%	0.62 (p=0.264)	Diclofenac	10	1996	0.50%	0.86 (p=0.414)
Ibuprofen	11	1985	0.55%	0.51 (p=0.073)	Ibuprofen	11	1985	0.55%	0.55 (p=0.450)

Reminder for comparison: The 6-month data as reported in the JAMA paper is reported as an annualized incidence of upper gastrointestinal tract complications (see page 1251 Figure 2A). In the JAMA paper, the NSAID data is not broken down

Six Month Data from the JAMA paper (annualized rate in person-years)

	# of events	Denominator (person-years)	Event rate	Relative Risk (confidence interval)
6 mo. Risk				
Celebrex	11	1441	0.76%	
NSAID	20	1384	1.45%	0.53 (0.26-1.11)

Questions to consider:

1. Can you calculate Number Needed to Harm from these tables?
2. Is it possible to compare the 6-month data as reported in JAMA to the 12-month data found on the FDA Web site (i.e. can you compare a rate with risk)?
3. What would you recommend to the editorial board of JAMA? To the FDA? To the pharmacy and therapeutics committee at your hospital? To your patient?

ANSWERS to HARM/RCT Example Teaching Package

Educational Package: Part I

Teaching Scenario:

Study Design	How you select your starting population of patients	Comparison group	Assessments and direction of inquiry	Strengths / Weaknesses
1) Case-Control	Identification of a group of people who already have the outcome you are interested in (CASES)	Identification of a group of people who do NOT have target outcome but are similar to cases in many ways (MATCHED CONTROLS)	Because the cases have already developed the outcome, you look BACK to determine the exposure status in each group.	Strength: Requires small sample size, studies can be done by looking back in time, thus, require little time to do. Weakness: Susceptible to bias, limited validity.
2) Cohort	Identification of exposed group (COHORT)	Identification of non-exposed group (COHORT)	Follows groups FORWARD in time monitoring for the occurrence of the outcome or adverse events.	Strength: Can be used when outcomes are harmful, and/or occur infrequently Weakness: Exposed and nonexposed groups may begin with different risk of harmful outcome (i.e. the groups may not be the same wrt confounding variables).
3) RCT	Random selection of groups that are balanced with respect to known and unknown determinants of outcome.	The comparison is Between Randomized Groups, one group with exposure and one group without the exposure.	FORWARD measurement of occurrence of the outcome or adverse events.	Strength: Less bias associated with the groups at the start. Weakness: may be unethical to randomize people to harmful interventions. May take large samples and years of study to capture rare and serious events

Although the paper that we found to answer our HARM question for Mr. G is a RCT, clinicians will often not find RCT to help determine HARM. RCTs are often not feasible if the outcome in question is truly harmful, posing an ethical challenge to randomization. Further, rare and serious events may only occur after following large numbers of people (e.g. tens of thousands of patients) for prolonged periods of time. Therefore, it is essential to gain an appreciation for the methodology, strengths and potential pitfalls (biases) associated with case-control and cohort studies. The following list of papers might be considered for teaching and learning about these methodologies:

Case-control study:

Phenylpropanolamine and the risk of hemorrhagic stroke. [NEJM 2000;343:1826-32.](#)

Prospective cohort study:

Incidence of and risk factors for ventilator-associated pneumonia in critically ill patients. [Ann Int Med. 1998;129:433-440.](#)

Population-based case-control study:

The risk of myocardial infarction associated with antihypertensive drug therapies. [JAMA 1995;274:620-625.](#)

Nested case-control study:

The relationship of postoperative delirium with psychoactive medications. [JAMA 1994;272:1518-1522.](#)

Critical Review Form for Harm

Citation: Silverstein FE. et. al. Gastrointestinal toxicity with celecoxib vs. nonsteroidal anti-inflammatory drugs for osteoarthritis and rheumatoid arthritis. [JAMA. 2000;284:1247-1255](#)

Are the Results Valid?	
Did experimental and control groups begin the study with a similar prognosis?	
Did the investigators demonstrate similarity with respect to all known determinants of outcome; did they adjust for differences in the analysis?	This trial was conducted as a randomized trial. Therefore, randomization, if done correctly should yield equal distribution of characteristics (both known and unknown) between the two groups. Specifically, baseline characteristics of the treatment groups were similar with respect to known risk factors for PUD including age, presence of a rheumatologic disorder, h/o GI bleed or ulcer, H. Pylori infection, and use of tobacco, alcohol, aspirin, steroids or anticoagulants.
For case-control studies: Were exposed patients equally likely to be identified in the two groups?	Not applicable as this is an RCT
Did experimental and control groups retain a similar prognosis after the study started?	
Were the outcomes and exposures measured in the same way in the groups being compared?	Yes. Patients were followed with prospective collection of information on utilization of study medications, adverse events, and upper GI complications. Assessors were blind to treatment group. A GI events committee reviewed each case of potential ulcer complication and assigned it as meeting or not meeting the defined criteria for upper GI complication.
Was follow-up sufficiently complete?	The authors report 100% follow-up at 6 months. However, this is misleading and based on a complicated definition of lost to follow-up that is not sound. Ideally, they would have followed everyone to six months irrespective of whether they were taking the medication and included all events in the analysis. Because a significant number of patients withdrew from the study. (1611 in Celecoxib group and 1784 in the NSAID group see Fig. 1 pg.1250) there is concern that loss to follow-up destroys the balance created by randomization, and leaves us with unbalanced groups. The loss to follow-up is, in fact, different in each group and we are not given prognostic information about those lost. Therefore, we cannot easily determine the direction of the bias (e.g. did the bias favor or work against the Celecoxib?) In the worst case scenario (if we assume everyone on Celecoxib has bad outcomes) any benefit of celecoxib would be lost.

<p>What are the Results?</p> <p>How strong is the association between exposure and outcome?</p> <p>How precise is the estimate of risk?</p>	<p>Data are reported as annualized incidence rates (# of events per 100 patient-years of exposure). Some patients were at risk for study endpoints for less than the full 6-month follow-up (e.g. after a patient had a study endpoint, they were censored or taken out of the pool of patients who could further contribute to the endpoints.)</p> <p>Primary outcome measure was the rate of ulcer complication alone. Secondary outcome measure was a combined endpoint of symptomatic ulcers or perforation, gastric outlet obstruction, and bleeding.</p> <p>Results Bottom line:</p> <p>Primary outcome of rate of ulcer complications alone were 0.76% for celecoxib vs. 1.45% for NSAIDs (p=0.09) – a trend that does not reach statistical significance.</p> <p>Secondary outcome of rates of combined ulcer / perforation / obstruction / bleeding were 2.08% (celecoxib) vs. 3.54% (NSAIDs; p=0.02). Thus, combined outcomes were statistically lower with celecoxib. (Abs risk difference: 1.46%, NNH=69)</p> <p>The importance of concurrent aspirin use can be underscored by comparison of the subgroups that were ASPIRIN users (20.6% of patients) vs. NON-ASPIRIN users.</p> <p>ASPIRIN users had a higher risk of ulcer complications and there was no difference between ulcer complications for celecoxib vs. NSAIDs in the ASPIRIN user group (2.01% vs. 2.12% p=0.49).</p> <p>NON-ASPIRIN users had lower ulcer complications in celecoxib vs. NSAIDs (0.43% vs. 1.27% p=0.04) and combined ulcer or ulcer combination (1.40% vs. 2.91% p=0.02)</p> <p>Table 3 summarizes the results of the GI events committee.</p> <table border="1" data-bbox="521 894 1377 1318"> <thead> <tr> <th></th> <th>Celecoxib</th> <th>NSAID</th> <th>RR (CI)</th> </tr> </thead> <tbody> <tr> <td>Upper GI bleed: (person-yrs)</td> <td>11/1441</td> <td>20/1384...</td> <td>0.53 (0.26-1.11)</td> </tr> <tr> <td>Ulcer or Ulcer complication: (person-yrs)</td> <td>32/1441</td> <td>51/1384...</td> <td>0.57</td> </tr> <tr> <td colspan="4">Subset who were NON-ASPIRIN Users:</td> </tr> <tr> <td>Upper GI bleed: (person-yrs)</td> <td>5/1143</td> <td>14/1101...</td> <td>0.35 (0.14-0.98)*</td> </tr> <tr> <td>Ulcer or Ulcer complication: (person-yrs)</td> <td>16/1143</td> <td>32/1101...</td> <td>0.48 *</td> </tr> </tbody> </table> <p>*p<0.05</p> <p>Table 4 summarizes additional adverse events. There was a lower incidence of reductions in hematocrit in the Celecoxib group as well as lower adverse renal events and HTN.</p>		Celecoxib	NSAID	RR (CI)	Upper GI bleed: (person-yrs)	11/1441	20/1384...	0.53 (0.26-1.11)	Ulcer or Ulcer complication: (person-yrs)	32/1441	51/1384...	0.57	Subset who were NON-ASPIRIN Users:				Upper GI bleed: (person-yrs)	5/1143	14/1101...	0.35 (0.14-0.98)*	Ulcer or Ulcer complication: (person-yrs)	16/1143	32/1101...	0.48 *
	Celecoxib	NSAID	RR (CI)																						
Upper GI bleed: (person-yrs)	11/1441	20/1384...	0.53 (0.26-1.11)																						
Ulcer or Ulcer complication: (person-yrs)	32/1441	51/1384...	0.57																						
Subset who were NON-ASPIRIN Users:																									
Upper GI bleed: (person-yrs)	5/1143	14/1101...	0.35 (0.14-0.98)*																						
Ulcer or Ulcer complication: (person-yrs)	16/1143	32/1101...	0.48 *																						
<p>How can I apply the results to my patient care?</p>																									
<p>Were the study patients similar to patients in my practice?</p>	<p>The patients are outpatients from 386 centers in Canada and the US. Review of Table 2 reveals an older population (mean age = 60), white (nearly 90%) with longstanding OA or RA. These patients are likely similar to my patients.</p>																								
<p>Was the duration of follow-up adequate?</p>	<p>6 months is shorter than the duration of therapy for many patients. Although there are certainly situations in which trials of anti-inflammatory medications can be short (e.g. in the case of acute injury), for the majority of patients with OA and RA medications are continued indefinitely.</p>																								

<p>Should I attempt to stop the exposure?</p>	<p>The dose of celecoxib used in this study is far greater than would be used in clinical practice. This may over-estimate risk of toxicity, however, the limited follow-up time of 6 months does not answer the question of ultimate safety in patients who are on these medications indefinitely.</p> <p>The question here is really should we substitute one drug for another. From these data, it seems that there might be some benefit to using celecoxib in preference to ibuprofen. How much better and whether it is worth the considerable extra cost, is a matter of balance of priorities for the individual doctor / patient pair. The Number Needed to Harm of 68 means that we would have to treat 68 patients with NSAID (compared to celecoxib) to cause one extra combined endpoint of ulcer/perforation/obstruction/bleed. This would be in line with other therapies. However, if the cost of medication is prohibitive, this may not be an option for our individual patient.</p>
---	--

Adapted from McMaster Evidence-based Clinical Practice Workshops

Critical Review Form for Therapy

Citation: Silverstein FE. et. al. Gastrointestinal toxicity with celecoxib vs. nonsteroidal anti-inflammatory drugs for osteoarthritis and rheumatoid arthritis. [JAMA. 2000;284:1247-1255](#)

Are the Results Valid?	
Did experimental and control groups begin the study with a similar prognosis?	
Were patients randomized?	Yes
Was randomization concealed?	Yes. The patient assignment was done by an interactive voice response system according to a computer generated, randomization schedule.
Were patients analyzed in the groups to which they were randomized?	Statistical analyzes were conducted on the intent-to-treat population defined a priori in the protocol as consisting of all patients who received at least 1 dose of assigned study medication.
Were patients in the treatment and control groups similar with respect to known prognostic factors?	Yes. They were similar with respect to demographic characteristics, risk factors for GI bleeding, H. Pylori infection, medication usage (aspirin, steroids, and anticoagulants) and comorbidity.
Did experimental and control groups retain a similar prognosis after the study started?	
Were 5 important groups (patients, caregivers, collectors of outcome data, adjudicators of outcome, data analysts) aware of group allocation?	Blinding is mentioned to patients, care providers, assessors of the clinical outcomes and drug company personnel.
Aside from the experimental intervention, were groups treated equally?	Yes.
Was follow-up complete?	See HARM appraisal sheets
What are the Results? How large was the treatment effect? How precise was the treatment effect?	See HARM appraisal sheets
How can I apply the results to my patient care?	
Were the study patients similar to my patient?	See HARM appraisal sheets
Were all patient-important outcomes considered?	See HARM appraisal sheets
Are the likely benefits worth the potential harms and costs?	See HARM appraisal sheets

Adapted from McMaster Evidence-based Clinical Practice Workshops

Educational Package: Part II

The rest of the story

Making sense of the additional data: Number Needed to Treat

Six Month Data from the FDA Web site [remember this is 6-month risk, NOT annualized (person-yrs) as in the JAMA paper]

	# of events	Denominator (# of people)	Event rate	Relative Risk	Absolute Risk Difference	Number Needed to Harm
6 mo. Risk						
Celebrex	11	3987	0.28%			
NSAID	20	3981	0.50%	0.56 (p=0.092)	0.22% or 0.0022	455
NSAID Data from 6 months is broken into the 2 groups below:						
Diclofenac	9	1996	0.45%	0.62 (p=0.264)	0.17% or 0.0017	588 (Not significant)
Ibuprofen	11	1985	0.55%	0.51 (p=0.073)	0.23% or 0.0023	435 (Not significant)

Twelve Month Data from the FDA Web site, uncensored

	# of events	Denominator (# of people)	Event rate	Relative Risk	Absolute Risk Difference	Number Needed to Harm
12 mo. Risk						
Celebrex	17	3987	0.43%			
NSAID	21	3981	0.53%	0.81 (p=0.640)	0.10% or 0.001	1000 (Not significant)
NSAID Data from 12 months is broken into the 2 groups below:						
Diclofenac	10	1996	0.50%	0.86 (p=0.414)	0.07% or 0.0007	1429 (Not significant)
Ibuprofen	11	1985	0.55%	0.55 (p=0.450)	0.212% or 0.0012	833 (Not significant)

Reminder for comparison: Six Month Data from the JAMA paper [annualized (person-yrs)]

	# of events	Denominator (person-years)	Event rate	Relative Risk (confidence interval)	Absolute Risk Difference	Number Needed to Harm
6 mo. Risk						
Celebrex	11	1441	0.76%			
NSAID	20	1384	1.45%	0.53 (0.26-1.11)	0.69% or 0.0069	145 persons-years

Educational Package Part II:

The rest of the story

Making sense of the additional data:

Overview:

Because the data in the JAMA paper and the data on the FDA website are not presented in a uniform, consistent way, it may be hard to pull the information together and come to definitive conclusions. However, the presence of additional data that was not reported must shake our faith in the data that is reported. At a minimum, we must look with a cautious eye to consider whether the comparative benefit of celecoxib compared to ibuprofen and diclofenac is supported by the full body of data that includes additional adverse events in the months following 6 months.

From a teaching point of view, there are several 'teachable moments' that come out of this real life example of biased reporting. The following are some of these teaching opportunities

1. Risk vs. Rate data:

The data on the FDA website does not give us information on annualized rates (person-years.). This creates the opportunity for discussion on several points:

- What is the difference between risk and rate?

In the JAMA paper, the authors report the absolute numbers of events in the entire sample over the 6-month period in Table 3. Therefore, we could calculate the risk of events in the population over the entire 6-month period of time. This does not give us any information on WHEN the events happened. The time to event analysis and annualized incidence takes into account how long the study subjects were on the medication prior to an adverse event.

- Can we compare risk vs. rate data?

Because we do not have the one-year data in person-years, it is difficult, if not impossible, to compare the annualized rate over 6 months to the 12-month risk data. We know that many patients did not contribute to the outcomes for the entire 6-month or 12-month period of time. Risk calculated from this data does not reflect true risk because they stopped following some people (i.e. they were censored). Although we can calculate the number needed to harm (NNH) in each situation, we cannot make a clean comparison for the risk vs. rate data.

2. The Ethics of drug company sponsored research:

It is notable that all of the authors were either Pharmacia employees or paid consultants. There were NO unpaid authors. This is clearly disclosed at the end of the article in 'financial disclosures'. How should we view this information as we read an article in the future? Of note, there has been no follow-up report in JAMA or clarification in the published literature to bring to public attention the additional, unreported data.

Returning to the patient:

Finally, we must decide what to tell Mr. G! On the basis of the available information, there is not strong evidence to support the extra expense to him, given his limited income. Thus, we switched him to ibuprofen!

Harm Sample Packet #2: Post-Op Delirium - Case-Control Study

Learning Objectives:

After working through this packet, you should be able to:

1. Apply the validity criteria for harm studies to an article concerning a question of adverse events due to pharmacotherapy.
2. Understand the methodology, strengths, and weaknesses of a case-control study.
3. Understand the meaning of an odds ratio and how to apply it to a clinical case.
4. Understand the principle of confounding and be able to generate examples of possible confounders in this example.

Assessment of the Case:

You are asked to consult on a patient on the orthopedics inpatient ward. The patient is a 69-year-old Caucasian woman who is s/p total knee replacement 3 days ago. She is an active 69-year-old who lived independently prior to her hospital stay. Her high level of activity (which had included a 2-mile walk three times each week) was recently halted due to her worsening osteoarthritis. Her past medical history includes hypertension and hypercholesterolemia.

Her surgery was uncomplicated, however, she had considerable post-operative pain and anxiety about being hospitalized and away from home. On postoperative day #3, she became agitated, pulled out all of her intravenous lines and complained of seeing insects crawling around her room. She became abusive to hospital staff and accused the nurses of trying to poison her with her medications. A search for infection, ischemia, and metabolic abnormalities that might explain her delirium is not fruitful.

Her current medications include HCTZ 25 mg po qd, simvastatin 20 mg po qd, lorazepam 1-2 mg i.v. q2 prn agitation and Benadryl 25 mg po qhs prn sleep. She is also getting the patient controlled administration of iv morphine.

The orthopedic surgery resident wonders if any of her medications might be contributing to her delirium. Specifically, she asks you whether benzodiazepines cause postoperative delirium. She is also concerned that the morphine PCA may be contributing to the patient's delirium. However, she feels strongly that she should treat the pain aggressively and would like to know whether it is safe to continue this medication.

Harm Sample Packet #2:

Post-Op Delirium - Case-Control Study

Asking the clinical question(s):

You frame the following clinical questions:

- ? In postoperative patients, are benzodiazepines associated with delirium?
- ? In postoperative patients, are narcotic drugs (including morphine) associated with delirium?

You are in quite a rush, so you log onto your hospital's library website and decide to search for an evidence summary in ACP Journal Club. You enter the following search strategy:

Search History:	Results
1. Postoperative	134
2. Benzodiazepines	21
3. 1 and 2	1

The single abstract that comes up is entitled Meperidine and benzodiazepines were associated with postoperative delirium. ACP Journal Club volume 122 p 80, May-June 1995.

Within a few seconds, you pull up the page long article summary in ACP journal club. The summary does seem to address your clinical questions, specifically discussing benzodiazepines and narcotics in postoperative patients. In reading the summary, you note that it is a 'nested case-control study' and you quickly acknowledge (to yourself) that you are not fully comfortable with this methodology. You decide that these are important clinical questions that you will face repeatedly as a consultant on the surgical service. Therefore, you decide to print the original article. Fortunately, this is available online in full text. You print the paper (as well as the ACP journal club review) for your review this evening.

Readings and Citation:

ACP Journal Club summary: Meperidine and benzodiazepines were associated with postoperative delirium. ACP Journal Club. [V122:p80, May-June, 1995.](#)

Original Reference:

Marcantonio ER, Juarez, G, Goldman, L et al. The relationship of postoperative delirium with psychoactive medications. [JAMA 272\(19\);1994:1518-22.](#)

Background reading:

Users' Guide to the Medical Literature:
Therapy and Harm: An Introduction
Harm Section

ANSWERS CRITICAL REVIEW FORM FOR HARM – Example #2 - Case-Control Study

Citation: The relationship of postoperative delirium with psychoactive medications. JAMA. 1994; 272:1518-1522.

Are the Results Valid?	
Did experimental and control groups begin the study with a similar prognosis?	
<p>Did the investigators demonstrate similarity with respect to all known determinants of outcome; did they adjust for differences in the analysis?</p>	<p>Yes. This study is a case-control study that was nested within a prospective cohort. The cohort study was designed to derive a clinical prediction model for postoperative delirium. 1341 patients were identified at the time of admission and all patients underwent an extensive preoperative assessment including review of medical history, physical examination, functional and cognitive testing and laboratory tests.</p> <p><u>Cases</u> were patients who developed delirium on post-operative day 2→5. (Post op day #1 was excluded due to concern that anesthesia effect might mimic delirium).</p> <p><u>Controls</u> were matched to each case from the population of patients in the prospective cohort who did NOT develop delirium. Controls did not develop delirium but had the same preoperative risk for delirium as the case patient based on the seven criteria (listed below). Controls had to still be in the hospital on the day the delirium developed in the case patient. Up to 2 matched controls were identified for each case. If no control patient met the criteria for a perfect match, a control was selected with the closest preoperative risk for delirium. No control was matched to a case with >10% difference in calculated risk. All controls were used only once.</p> <p>2 matched controls were used for 63 of the 91 patients who developed delirium (69%). 1 matched control was used for the remaining. Thus, the total number of matched controls was 154.</p> <p><u>Delirium risk assessment:</u> All patients in the cohort were interviewed with respect to seven criteria pertaining to risk for delirium: age, poor cognitive function, poor physical function, self-reported alcohol abuse, markedly abnormal preoperative serum sodium, potassium or glucose levels, aortic aneurysm surgery and non-cardiac thoracic surgery.</p>
<p>For case-control studies: Were exposed patients equally likely to be identified in the two groups?</p>	<p>Yes. All 1341 study patients in the prospective cohort underwent identical daily structured interviews on postoperative days 2→5, or until the day before discharge if discharge occurred prior to day 6.</p> <p>Within this cohort, patients with delirium were identified if they met criteria for delirium using the <i>Confusion Assessment Method (CAM) diagnostic algorithm</i>.</p>
Did experimental and control groups retain a similar prognosis after the study started?	
<p>Were the outcomes and exposures measured in the same way in the groups being compared?</p>	<p>All risk and pre-operative data were collected prospectively on all patients, regardless of the ultimate development of delirium. Data on medication exposure was obtained by review of the nurses' medication flow sheet for those medications given (not simply those ordered) as well as from the pain service orders (which controlled the administration of epidural and patient-controlled analgesia).</p> <p>A reviewer blinded to the study hypothesis collected chart review data. For both cases and controls, medication exposures were recorded for the 24-hour period before the delirium developed. Care was taken to record medications given in the period prior to delirium to exclude medications given after signs of delirium developed.</p>

	<p>Possible confounders, specifically with respect to the use of benzodiazepines might be that the early signs of delirium (e.g. sleeplessness and agitation) might be treated with benzo's. Therefore, it is possible that cases might be 'exposed' to these drugs as a sign of early delirium that was not yet diagnosed, as opposed to benzodiazepines causing the delirium. Other confounders relate to the possibility that other coincident factors may have been related to the use of these medications (e.g. sleep deprivation, agitation, pain control).</p>
Was follow-up complete?	<p>As this was a nested case-control study in the context of prospective data collection. Follow-up was complete for those who remained hospitalized for the 5 days of the data collection period. Those who were discharged early were not followed outside the hospital.</p>
<p>What are the Results?</p> <p>How large was the treatment effect?</p> <p>How precise was the treatment effect?</p>	<p>Three medication classes were selected for study: narcotics, benzodiazepines, and anticholinergics.</p> <p><u>Narcotic Agents</u>: Only <u>Meperidine</u> was associated with Delirium (OR 2.7; 95% CI, 1.3 to 5.5). Of note, narcotic medications were commonly used with roughly 95% of all patients (cases and controls) receiving narcotics during the 24-hour period of analysis.</p> <p><u>Epidural anesthesia</u> was associated with delirium (OR 2.3; CI, 1.2 to 4.4). {This may be related to the high proportion of patients who had epidural anesthesia with meperidine as the agent used.}</p> <p><u>Patient-controlled anesthesia</u> was not associated with delirium (OR 1.1; CI 0.5 to 2.2)</p> <p><u>Benzodiazepines</u> These drugs were used less frequently with 21% of cases and 8% of controls receiving benzodiazepines during the 24-hour period of analysis.</p> <p>Overall, benzodiazepines were associated with delirium (OR 3.0; CI 1.3 to 6.8).</p> <p>Long-acting benzo (OR 5.4 CI, 1.0 to 29.2) > short acting (OR 2.6, CI 1.1 to 6.5)</p> <p>High-dose benzo (OR 3.3 CI 1.0 to 11) > low dose (OR 2.6 CI 0.8 TO 9.1)</p> <p><u>Anticholinergic agents</u> were not associated with delirium (OR 1.5 CI 0.6 to 3.4). {The relatively small proportion of patients receiving these drugs during the 24-hour period of analysis (9%) may have limited power to detect a difference in this category.}</p>
<p>How can I apply the results to my patient care?</p>	
Were the study patients similar to my patient?	<p>These patients were all surgical patients admitted to Brigham and Women's Hospital in Boston for major elective non-cardiac procedures. The average age of the patient was 73 years. Roughly half of the patients in the study underwent an orthopedic procedure. 50% were women and the participant's race is not reported. Thus, for those characteristics described, the patients in the study are quite similar to my own patient.</p>
Were all patient-important outcomes considered?	<p>Follow-up was 5 days postoperative or to discharge. This length of time is likely sufficient for this question, but you might miss cases who were discharged early or events that were late occurring.</p>

<p>Are the likely benefits worth the potential harms and costs?</p>	<p>We really considered two questions as a result of the orthopedic consult: Are benzodiazepines associated with postoperative delirium, and are narcotics, specifically PCA morphine associated with postoperative delirium. In this study, the answers should be looked at separately.</p> <p>Question 1: Are benzodiazepines associated with postoperative delirium? Yes. The odds ratio for development of delirium, when exposed to benzodiazepines, was 3.0. In addition, this paper gives evidence that the association may be stronger for longer acting medications as well as for higher dose of medications. In fact, for our patient, she is on a long acting drug (lorazepam) with potential for “high doses” (anything greater than 1 mg). It would be prudent to stop this exposure in this patient.</p> <p>Question 2: Are narcotics, specifically PCA morphine, associated with postoperative delirium? In this study, only meperidine was associated with delirium (OR 2.7). Morphine, by PCA (the regimen our patient was on), was not associated with delirium (OR 0.9; CI 0.4-1.9). However, one should maintain some caution and consider that it is possible that the lack of association was related to insufficient power to detect a statistical difference. In this setting, it might be reasonable to cautiously continue our patient’s PCA morphine pain management at the lowest possible dose, including a strategy to augment pain management with non-narcotic medications (e.g. NSAIDS) and switch to these medications when pain control allows.</p>
---	---

Adapted from McMaster Evidence-based Clinical Practice Workshops

Harm Sample Packet #3: Case Series/Report

A great case of lousy evidence. A look at a case series from the FDA.

Learning Objectives:

1. Consider the different types of methodology that are used to look at questions of Harm.
2. Understand the strengths (if there are any) and weaknesses of a case-control study
3. Discuss what you do when there is no controlled trial evidence for a clinical question. How do you apply this information to patient care?


Case Scenario:

The patient is a 15-year-old young man with severe acne. He presents after trying and failing over the counter and other prescribed therapies requesting Accutane. He states that his best friend tried this and it made his acne completely go away. He is very stressed over social issues at school. This school year has been very difficult for him and he feels that his acne has played a role in some of his social isolation. He is doing more poorly in school and his parents report that he is increasingly emotionally labile and irritable. When you ask his mother if she suspects that he is depressed, she answers, “no more than any other 15-year-old boy on the planet.”

You write a prescription for Accutane and the teenager happily leaves your clinic. Your medical student asks you whether there is an association between the use of Accutane and depression and suicide. You say that you are aware that there is a warning pertaining to this. You send your medical student to try to find an answer.

The **PICO** clinical question is defined as:

In Patients with Acne, is there a relationship between the Intervention Accutane (isotretinoin) and Outcome of depression or suicide?



Being rather a competitive individual, you decide to ‘race’ your medical student. You are going to look this up on UpToDate to see what it says about this relationship and she is going to do a quick Medline search. Much to your chagrin, you return at the exact same moment with the same citation. She however, has pulled the full-text pdf article off the web, whereas you only have the abstract from UpToDate.

Each of you found the following article:

Wysowski, DK, Pitts, M, Beitz, J An analysis of reports of depression and suicide in patients treated with isotretinoin. [J Am Acad Dermatol 2001;45:515-9.](#)

At first glance, you can see that this is not a controlled trial, but in your assessment, it is the best data available on this question. Please review the paper and decide whether the evidence presented should make you change your thinking about the patient you saw this morning.

ANSWERS for Case Series/Report Teaching Harm/Etiology Example #3

Clinical Question:

Does isotretinoin use cause depression and/or suicide in treated patients with acne?

Reference:

Wysocki DK, Pitts M et al. An analysis of reports of depression and suicide in patients treated with isotretinoin. *Journal of American Academy of Dermatology*, 45(4); 515-519, October 2001.

Methods

Design: Prospective case series from post-marketing voluntary FDA reports of adverse drug events, dating from 1982 to May 2000.

Setting: The United States, post-marketing FDA Adverse Event Reporting System Database.

Patient Population: US patients (presumably with severe acne) treated with isotretinoin, compared with all other licensed drugs in the US with voluntary drug reports during the same years.

Description of exposures being considered: Isotretinoin, at variable doses and duration of treatment compared with other drugs licensed in the US.

Analysis: No statistical analysis was performed. Numbers of reports of depression and suicide attempts were compared between all drugs in the FDA Adverse Event Reporting System database. Outcomes: Number of reports of depression and suicide attempt.

Follow-up: None.

Validity

Were the comparison groups clearly delineated and treated similarly outside of the putative exposure?

This is unknown. In this prospective series comparing patients treated with different classes of drugs (presumably with different disease states), one would assume they were not treated similarly. No data related to this were presented.

Did the groups have similar exposures to other possible determinants of the outcome?

This is also unknown. Since the FDA database contains patients treated with all drugs approved for marketing in the US, for different indications and diseases – presumably they were not treated similarly.

Were outcomes and exposures measured similarly in each comparison group?

Yes. All reports were voluntary post-marketing adverse event reports with diagnostic codes of suicide or suicide attempt, and depression and related disorders.

Was follow-up of sufficient duration?

Yes, all years of available data since approval of isotretinoin were included in the analysis.

Were all patients accounted for and analyzed?

Probably not. Voluntary reporting likely missed a significant number of cases. Data for isotretinoin may be significantly biased as 49% of cases were reported in 1998 – after suicide and depression were added to the product labeling in February 1998!

Is there a reasonable temporal relationship between exposure and outcome?

Yes. All patients were treated with isotretinoin prior to the diagnosis of depression and/or suicide (or attempt). Some patients had medication discontinued prior to diagnosis.

Is there a dose-response gradient?

This could not be addressed with this study design. Only patients with events (depression or suicide) were documented with the analyzed reports. Without a denominator (number of patients treated) and doses received, dose response could not be determined.

Do the study population characteristics describe your patient?

Yes.

Results What is the strength of the association between exposure and outcome?

This study did not permit a quantification of the risk. Questions of reporting bias (with the significant number of cases reported after labeling of the drug changes in 1998) severely limit the conclusions we can draw about the number of reports associated with isotretinoin. There were a relatively large number of reports of events for isotretinoin (the only nonpsychiatric drug in the top 10 drugs ranked for suicide attempts), and case reports of positive re-challenges with the drug.

How precise are the risk estimates for the exposure?

Cannot be determined from this study, as the design did not allow statistical analysis of this sort.

What is the absolute level of risk from this exposure?

This remains unknown for the same reasons as above. See the limitations below.

Comments

Strengths and Weaknesses of Study (internal and external validity)

This study provides the largest available series of patients reported with psychiatric complications that may be related to isotretinoin. Certainly this relatively systematic approach for monitoring reports of drug adverse events is superior to expert opinion or a simple case series, but it is limited by methodology, lack of an appropriate comparison group and reporting bias.

Should the study results cause avoidance of the exposure?

For a number of reasons, isotretinoin should be used with caution and close follow-up (teratogenicity, lipid abnormalities...), and should be reserved for patients with significant and/or refractory acne. These data in the context of previous literature suggest that close monitoring and follow-up of patients is clinically appropriate – but they do not necessarily suggest the drug should not be used.

Next steps for further study of this problem:

To better answer this question, a study with an appropriate, matched control group (such as a case-control study) or prospective cohort study with comparison to non-treated, matched controls could be conducted. Given the low frequency of significant events (suicide attempts, and diagnoses of depressive disorders) a large prospective study would be necessary and thus is unlikely to be undertaken. Careful follow-up and continued post-marketing reporting of events may yield further evidence of the risk – though it is likely we will not have a certain answer.

Returning to the Clinical Scenario:

In the meantime, despite the flaws in this study design, this represents the strongest piece of evidence available to date on this question. Therefore, we will have to consider this information (in the context of the design flaws) and still make a clinical decision. After consideration of the evidence, you do not feel that the strength of the evidence should stop you from a monitored trial of isotretinoin in the 15-year-old patient you saw in the morning and you do not change your recommendation. You do however, reiterate your commitment for close follow-up.

Teaching points to highlight for this paper:

This paper highlights the unfortunate reality that we frequently face when searching real life questions in the course of caring for our patients. It is not uncommon that the level of evidence that we seek simply does not exist. Therefore, like it or not, we must make clinical decisions with the best information that is available to us. Discussion of this paper might lead to a perfect opportunity to discuss the meaning of 'the hierarchy of evidence'. In addition, this paper can be a jumping point for discussion of the methodologic stumbling blocks that face researchers when they are attempting to address issues of harm. One might take the opportunity to review different methodologies in the context of how a researcher might design different kinds of studies to answer this very question and the pros and cons of each method.

Prognosis

Prognosis Core Concepts

Notes on learning and teaching about **Prognosis**:

1. **Background:** The concept of prognosis (the estimated probability of outcomes of disease processes over time) is at the core of what clinicians do. Cohort studies typically follow one or more groups of people with similar characteristics over time and provide information on the occurrence of outcomes. Cohorts can be prospective or retrospective; the subjects can be population or community-based or tertiary care patient populations. Understanding how and when the cohort is defined and how and when the outcomes are measured is core to understanding the literature pertaining to prognosis.
2. Key concepts and terms
 - Cohort Study
 - Appropriate or inappropriate measurement of exposure
 - Statistical adjustment for prognostic imbalance
 - Lost to follow up
 - Risk factor vs. Prognostic factor
 - Bias
 - Incidence vs. Prevalence
 - Odds Ratio vs. Risk Ratio
 - Survival analysis (sometimes called Kaplan-Meier Analysis) and hazard ratio

Prognosis Critical Appraisal Form

Citation:

How serious is the risk of bias?	
Was the sample of patients' representative?	
Were patients classified into prognostically homogeneous groups?	
Was follow-up sufficiently complete?	
Were outcome criteria objective and unbiased?	
What are the results?	
How likely are the outcomes over time?	
How precise are the estimates of likelihood?	
How can I apply the results to patient care?	
Were the study patients and their management similar to those in my practice?	
Was the follow-up sufficiently long?	
Can I use the results in the management patients in my practice?	

Adapted from McMaster Evidence-based Clinical Practice Workshops and Users' Guide to the Medical Literature 3rd Ed.

Prognosis Teaching Package #1: Hepatitis A- A deadly partner?

New Patient Evaluation

Mr. W is a 43-year-old black male who is new to GMC. He has a remote history of a few 'experiments' with IVDU and was dx with Hepatitis C (by antibody and viral RNA titer) six months prior to your clinic visit. The diagnosis was made when he was donating blood. He had mild elevations of his transaminases, however, had a great function of his liver (PT-INR is normal). He underwent a liver biopsy which did not reveal any fibrosis and was started on a combination of interferon and ribavirin in GI clinic.

Since starting interferon/ribavirin, he has noted flu-like symptoms, worsening of depressive symptoms and general fatigue. These symptoms have resulted in some loss of work time for him as he works as a fork-lift operator and cannot work safely when he does not feel well. He controls his symptoms with Tylenol with some success. In addition, he has anorexia and has lost 15 lbs. since starting the therapy.

Medications: Paxil, lansoprazole

NKDA

Social history: married to a longstanding partner with whom he is monogamous. He has 6 children and 8 grandchildren and lives in South Boston. Nonsmoker- former drinker (none for the past 8 yrs.). No other recreational drugs. He is religious and gets much support from his church community.

PE: unremarkable

Labs sign: for WBC: 4.3, HCT: 40.8, PLT 228. You note that his HCT is dropping He is also Hep B SAB+

The patient is quite motivated to do anything he can do to impact the course of his disease, therefore during your visit, you focus on encouraging compliance with his interferon/ribavirin therapy and also you discuss other aspects of prevention. You note that he has already been exposed to Hep B but would like to check labs to determine whether he has been exposed to Hep A. He is somewhat concerned about this because he believes that his wife was told that she had been exposed to Hep A but not hep B or hep C.

In general, you would like to review data on prevention and health maintenance in the setting of known hep C infection. Specifically, you would like to know the data concerning superinfection with Hep A, in someone who has hep C.

You are very pushed for time. Please consider the following questions:

1. What are the possible resources you can use to answer some of these questions?
2. Will you use different resources to answer different kinds of questions that you have?
3. Specifically, consider how the information you find will help to change your management of your patient or your advise / counseling to him or his wife.

Paper(s) that will be read at the next session are:

Vento et al "Fulminant hepatitis associated with hepatitis A virus superinfection in patients with chronic hepatitis C." [NEJM 1998;338:286-90](#)

Questions	Article
<p>A. ARE THE RESULTS VALID?</p> <p>Sample of Patients? Was there a representative and well-defined sample of patients at a similar point in the course of the disease?</p>	<p>University of Verona, Italy. Prospectively followed 595 patients from June 1990-July 1997 (405 men / 190 women) with histological evidence of hep B or hep C. Patients were 'all' patients followed at that institution for at least 18 months.</p> <p>432—Hep C (antibody and RNA) 163—Hep B Sag + Hep B DNA in serum</p> <p>Comparison Cohort: 191 consecutive patients with acute hep A who were negative for hep C or hep B and NO underlying liver disease</p> <p>All 595 patients had staging liver bx between June 89 and May 90</p> <p>Table 1: p 287 shows considerable variability of hep B and hep C disease, however, ALL patients are naïve to hep A.</p> <p>Note: Inception Cohort: early and uniform—not necessarily in this group</p>
<p>Follow-up: Was follow-up sufficiently long and complete?</p>	<p>All patients followed for a minimum of 18 months. They do not appear to have lost anyone.</p>
<p>Outcome criteria: Were objective and unbiased outcome criteria used?</p>	<p>It seems that all patients had liver bx at the start of the study and also the following hep A (if they survived)</p>
<p>Prognostic factors: Was there adjustment for important prognostic factors?</p>	<p>They did not really adjust for other prognostic factors, but they did report several including:</p> <ol style="list-style-type: none"> 1. They did have liver bx prior to the onset of hep A, thus, they could account for pre-existing liver disease. 2. They also measured HLA phenotypes, levels of serum autoimmune studies such as ANA, anti-smooth muscle antibodies, and anti-asialoglycoprotein receptor antibodies.
<p>B. WHAT ARE THE RESULTS? How large is the likelihood of the outcome event(s) in a specified period of time?</p>	<p>hep B and HAV 10 / 163; fulminant hepatitis in 0/10; single pt with marked cholestasis. hep C and HAV 17/ 432; fulminant hepatitis in 7/17; 6/7 patients died</p>
<p>How precise are the estimates of likelihood?</p>	<p>There is no estimate of precision given. However, the numbers are very small for each individual group, thus, the confidence limits would be wide.</p>
<p>C. WILL THE RESULTS HELP ME IN CARING FOR MY PATIENTS? Were the study patients similar to my own?</p>	<p>We cannot easily tell from the text which patients came down with hep A and what was their extent of liver disease prior. These patients, followed in a liver clinic, may be sicker than our population.</p>
<p>Will the results lead directly to selecting or avoiding therapy?</p>	<p>The issue at hand is the prevention of hep A with vaccination. The suggestion in this paper does support vaccination in patients with hepatitis C infection.</p>
<p>Are the results useful for reassuring or counseling patients?</p>	<p>Yes. They will be useful in counseling patients to obtain vaccination.</p>

Adapted from McMaster Evidence-based Clinical Practice Workshops

Prognosis Example #3

Risk: The Bullwinkle Family in Brazil- Café con Azúcar

Learning Objectives:

To consider a question of prognosis/risk

Specific objectives:

- Clinical question formation
- Translation of the question to an effective search
- Critical appraisal of a cohort study
- Application of the evidence to the case

Vocabulary

- Incidence
- Prevalence
- Relative Risk
- Confounding variables
- Bias
- Stratified analysis
- Causality

Specific Questions to think about when you read

Validity Section

- Table I: Baseline characteristics—are the groups equal at baseline? Why or why not?
- Table I: For each characteristic that is NOT similar between groups, how would you expect it to impact the outcome (i.e. direction of bias)

Results Section

- Table II: What's a person-year?
- Table II: adjusted analyzes- what's that?
- What two different ways can you discuss precision or our confidence in the results?

Clinical Scenario:

Bullwinkle Family Vacation:

You are the personal physician for Mr. and Mrs. Bullwinkle who are traveling on a six-month tour around the globe. Their first stop is Brazil. While visiting the capital city, Brasilia, they pass by the building that is home to the Organización Internacional del Café. Posted on the door is a press release in English and in Spanish entitled "Coffee Drinkers at Lower Risk for Type 2 Diabetes."

Mr. Bullwinkle is a 59-year-old veterinarian who earned millions through the creation of an empire in the pet psychotherapy industry. He has a personal history of GERD and hyperlipidemia. Current medications include simvastatin and omeprazole. His BMI is 30. He does not smoke. He does drink alcohol (1-2 glasses of red wine each night with dinner "to protect his heart"). His second cousin twice removed, has diabetes. Mr. Bullwinkle is very health conscious and believes in prevention. He is concerned about diabetes because of the misery that it has caused his second cousin and immediately sends you an e-mail message from his wireless Blackberry to inquire as to how many cups of coffee he should drink each day to protect him from the development of diabetes. Currently, he does not drink any coffee but prefers tea. He says he will 'learn to love it' for the sake of his health but will likely have

to put in a lot of sugar in order to be able to tolerate the strong taste. He does not like artificial sweetener and has concerns about its healthfulness.

Knowing that you are a busy clinician, he understands that you may not answer him today but will be looking for the answer tomorrow. He is already locating all-day coffee shops and has listed his name on the mailing list for the Organización Internacional del Café so that they can ship him the necessary supply of coffee beans.

Clinical Question Formation:

PICO

Patient Population:

Intervention / Exposure / Prognostic Factor:

Comparison:

Outcomes:

Type of Question / Type of Study desired:

Searching:

You may select up to 3 search terms or concepts.

1. _____
2. _____
3. _____

Teaching Strategies to consider:

Workshop settings: To engage active learning in a workshop setting, you can divide the learners into small groups, each responsible for presenting a set of issues to the group. You can use file cards for giving out assignments as follows- to be presented in the following order:

1. Group 1: Clinical Question formation and search term choices (limited to 3 concepts). Also, group 1 will need to consider how the question might be answered by each of the following 3 study designs: Randomized controlled trial, cohort, and case-controlled study and the pluses and minuses of each.
2. Group 2: Define and discuss incidence, prevalence, confounding variables and bias. Give general definitions as well as what role these concepts play in this paper.
3. Group 3: Define and discuss relative risk, confidence intervals, adjusted analysis and confidence intervals.
4. Group 4: Application and discussion of causality

Are the Results Valid?	
Was the sample of patients' representative?	
Were patients sufficiently homogeneous with respect to prognostic risk?	
Was follow-up complete?	
Were objective and unbiased outcome criteria used?	
What are the results?	
How likely are the outcomes over time?	
How precise are the estimates of likelihood?	
How can I apply the results to patient care?	
Were the study patients and their management similar to my own?	
Was the follow-up sufficiently long?	
Can I use the results in managing patients in my practice?	

Adapted from McMaster Evidence-based Clinical Practice Workshops

Prognosis Example Coffee ANSWERS

Risk: The Bullwinkle Family in Brazil- Café con Azúcar

Learning Objectives:

To consider a question of prognosis/risk

Specific objectives:

- Clinical question formation
- Translation of the question to an effective search
- Critical appraisal of a cohort study
- Application of the evidence to the case

Vocabulary

- **Incidence:** Number of new cases of disease occurring during a specified period of time; expressed as a percentage of number of people at risk
- **Prevalence:** Proportion of persons affected with a particular disease at a specified time. Prevalence can be used to assess pre-test probability if it is assessed from a study of strong methodology
- **Relative Risk:** Also known as Risk Ratio

Outcome			
	Outcome Present	Outcome Absent	
Treated/ Exposed (Y)	a Outcome present in treated patient	b Outcome absent in treated patient	Y= Risk of Outcome in Treated Group =a/(a+b)
Control/ Not exposed (X)	c Outcome present in control patient	d Outcome absent in control patient	X= Risk of Outcome in Control Group =c/(c+d)

Risk Ratio:

The ratio of risk of outcome in treated/exposed group (Y) as compared with control/ not exposed group (X)

$$RR=Y/X = a/(a+b) / c (c+ d)$$

This always tells us whether the observed outcome (effect) occurs more or less often in the exposed group than in the unexposed group. Calculations for RR are identical whether you are asking a question about therapy or a question about Harm.

- **Confounding variables:** A factor (besides the one you are studying) that distorts the true relationship of the study variable of interest to the outcome. Confounders are likely to be unequally distributed in non-random methodologies, therefore, may introduce bias (see below).
- **Bias:** Systematic differences between groups which may skew the results leading to a deviation from the true result.

- **Stratified analysis:** “arranged in layers or strata” -- regarding analysis, it refers to pre-sorting of data by characteristics prior to running the analysis. When discussed in terms of randomization, it means pre-sorting data prior to randomization.
- **Causality:** To interpret whether the association is likely to be causal, consider the following factors: 1) strength / magnitude of the relationship; 2) dose-response relationship; 3) temporal relationship; 4) consistency to other studies; 5) reversibility; 6) biological plausibility.

Specific Questions and Tasks

Validity Section

- Table I: Baseline characteristics—are the groups equal at baseline? Why or why not?
- Table I: For each characteristic that is NOT similar between groups, how would you expect it to impact the outcome (i.e. direction of bias)

Results Section

- Table II: What’s a person-year
- Table II: adjusted analyzes- what’s that?
- What two different ways can you discuss precision or our confidence in the results?

PICO

Patient Population: Patients without known diabetes. This patient is a 59-year-old male with GERD and hyperlipidemia

Intervention / Exposure / Prognostic Factor: Coffee Consumption

Comparison: no coffee consumption

Outcomes: incidence of diabetes

Type of Question / Type of Study desired: Risk / Prognosis: Prospective Cohort

Searching:

"Coffee"[MeSH]	2605
"Diabetes Mellitus"[MeSH]	165,952
#1 AND #2	56
"cohort studies"[MeSH] OR "prospective studies"[MeSH]	482,497
#3 AND #4	9

Critical Review Form for Prognosis **ANSWERS**

Citation: Coffee Consumption and Risk for Type 2 Diabetes Mellitus. [Ann Int Med 2004;140:1-8.](#)

Clinical Question: Is coffee drinking associated with decrease in risk for Type 2 Diabetes Mellitus

Are the Results Valid?	
Was the sample of patients' representative?	<p>The patients were taken from 2 prospective cohort studies: Both studies included health professionals participating in a mailed survey questionnaire providing detailed information about their medical history, lifestyle and risk factors. Both studies used a methodology that included a collection of data at 2 yearly intervals through a mailed survey to reassess risk factors and also to update information on newly diagnosed diseases.</p> <p>For this analysis: exclusion criteria for use of patient information in this study were known baseline history of DM2, CAD or cancer, non-completion of a portion of the survey items, extremes of caloric intake (<800 kcal/day or >4200 kcal/day for men; <500 kcal or >3500 kcal for women).</p> <p>The Health Professionals Follow up-Study (HPFS): Established 1886: 51,529 male health professionals (dentists, optometrists, veterinarians, osteopathic physicians, podiatrists, and pharmacists) 40-75 years of age. Failure to complete >70 of 131 items on food questionnaire → exclusion. After exclusions, the analysis included 41,934 men.</p> <p>The Nurses Health Study (NHS) Established 1976: 121,700 female nurses 30-55 yrs of age from 11 states. Semi-quantitative food frequency questionnaire added in 1980. Failure to complete >10 items on food questionnaire → exclusion. After exclusions, analysis included 84,276 women.</p>
Were patients sufficiently homogeneous with respect to prognostic risk?	These were population-based prospective cohorts. The exclusion of preexisting disease was intended to take out prevalent disease at the time of entry to the study.
Was follow-up complete?	It is not clear what proportion of individuals were lost to follow-up in either cohort.
Were objective and unbiased outcome criteria used?	<p>Assessment of Coffee / Caffeine Intake: Validated dietary questionnaires sent to HPFS and NHS participants multiple times. Question asked: how often on average during the previous year did you consume coffee and tea? Both groups also added a second question about decaffeinated coffee and different types of caffeinated items (added in 1986 in HPFS and 1984 in NHS).</p> <p>Intake of caffeine was calculated by summing the caffeine content for a specific amount of each food during the prior year.</p> <p>A validation subset confirmed high correlations with 1-week diet diaries.</p> <p>Coffee consumption was categorized into 5 groups (never, <1cup per day, 1-3 cups per day, 4-5 cups per day, 6+ cups per day)</p> <p>Caffeine intake: categorized by quintiles.</p> <p>Assessment of Diabetes Cases (consistent with National Diabetes Data Group): New cases of DM defined by 1) symptoms + elevated fasting glucose levels \geq 140 or random measured glucose \geq 200; 2) 2+ elevated plasma glucose concentrations (\geq 200) on different occasions without symptoms 2+ hours after Glucose tolerance test; 3) use of insulin or oral hypoglycemics.</p>

	A validation subset confirmed the validity of this diagnostic algorithm of determining new dx of DM. Dx of Type II DM was confirmed by medical records in 98% of participants.																					
What are the results?																						
How likely are the outcomes over time?	<p>Baseline characteristics: (Table I)</p> <p>Higher coffee consumption was strongly associated with cigarette smoking and alcohol use. Also, higher coffee consumption was associated with total and saturated fats and magnesium and inversely associated with physical activity, intake of cereal fiber, glycemic load and tea.</p> <p>Coffee consumption was NOT associated with BMI.</p> <p>Development of DM:</p> <p>Overall, 1,333/4,1934 men (3%) and 4,085/84,276 women (4.8%) were diagnosed with DM.</p> <p>Relative Risk of Diabetes: according to coffee consumption (Table 2 pg 4)</p> <p>Multivariate Model: adjusted for age, BMI, physical activity, Family history of DM, hormonal use (women), tobacco, alcohol, total calories, quintiles of trans fat, glycemic load, cereal fiber, magnesium.</p> <table border="1" data-bbox="578 831 1365 1058"> <thead> <tr> <th>CUPS/ DAY</th> <th>MEN (CI)</th> <th>WOMEN (CI)</th> </tr> </thead> <tbody> <tr> <td>NEVER</td> <td>1.00</td> <td>1.00</td> </tr> <tr> <td><1</td> <td>0.98 (0.84-1.15)</td> <td>1.16 (1.05-1.29)</td> </tr> <tr> <td>1-3</td> <td>0.93 (0.80-1.08)</td> <td>0.99 (0.90-1.08)</td> </tr> <tr> <td>4-5</td> <td>0.71 (0.53-0.94)</td> <td>0.70 (0.60-0.82)</td> </tr> <tr> <td>6+</td> <td>0.46 (0.26-0.82)</td> <td>0.71 (0.56-0.89)</td> </tr> <tr> <td>P VALUE FOR TREND</td> <td>0.007</td> <td><0.001</td> </tr> </tbody> </table> <p>Relative Risk of Diabetes: according to Tea consumption (Table 2 pg 4) No statistically sign relationship (P value for trend: >0.2)</p> <p>Note: alternative analysis, which adjusted only for age and BMI, showed similar results.</p> <p>Relative Risk of Diabetes: according to caffeine intake (Table 3 pg 5) Statistically significant inverse association (P value for trends in the multivariate model: <0.001 in HPFS and NHS)</p> <p>Controlling for confounders: Multivariate analysis (as above)</p> <p>A modest inverse association was noted between decaffeinated coffee and risk of DM (table 2)</p> <p>Stratified analysis by BMI, Smoking status, Physical activity (table 4) suggests inverse association was independent of lifestyle risk factors.</p>	CUPS/ DAY	MEN (CI)	WOMEN (CI)	NEVER	1.00	1.00	<1	0.98 (0.84-1.15)	1.16 (1.05-1.29)	1-3	0.93 (0.80-1.08)	0.99 (0.90-1.08)	4-5	0.71 (0.53-0.94)	0.70 (0.60-0.82)	6+	0.46 (0.26-0.82)	0.71 (0.56-0.89)	P VALUE FOR TREND	0.007	<0.001
CUPS/ DAY	MEN (CI)	WOMEN (CI)																				
NEVER	1.00	1.00																				
<1	0.98 (0.84-1.15)	1.16 (1.05-1.29)																				
1-3	0.93 (0.80-1.08)	0.99 (0.90-1.08)																				
4-5	0.71 (0.53-0.94)	0.70 (0.60-0.82)																				
6+	0.46 (0.26-0.82)	0.71 (0.56-0.89)																				
P VALUE FOR TREND	0.007	<0.001																				
How precise are the estimates of likelihood?	<p>If we wish to apply this to a member of the general population without comorbid disease who wishes to drink coffee, then the population is appropriate. If, however, you wish to apply this to a patient in my general medicine clinic (who is very likely to have comorbid illness already, it may be less applicable.</p> <p>For Mr. Bullwinkle, the study population is directly relevant and he would have met inclusion criteria.</p>																					

How can I apply the results to patient care?	
Were the study patients and their management similar to my own?	Follow-up was up to 12 years in the HPFS study (1986-January 98) and 18 years in the NHS (1980-98). Follow-up was calculated from return of baseline information (in 1986 for men and 1980 for women) to the diagnosis of DM, death or end of follow-up period in 1998.
Was the follow-up sufficiently long?	Follow-up was up to 12 years in the HPFS study (1986-January 98) and 18 years in the NHS (1980-98). Follow-up was calculated from return of baseline information (in 1986 for men and 1980 for women) to the diagnosis of DM, death or end of follow-up period in 1998.
Can I use the results in managing patients in my practice?	<p>Caffeine intake is one of many lifestyle issues that physicians counsel patients as well as family members about. The data presented from these two studies supports an inverse association between caffeine intake and diabetes incidence. However, these observational data cannot prove a cause-effect relationship.</p> <p>For Mr. Bullwinkle, we can counsel him to drink coffee while in Brazil if he enjoys doing that. However, the coffee may also upset his GERD and without stronger evidence that this will decrease his risk of development of diabetes, I would encourage him to limit his coffee drinking if it worsens his symptoms.</p> <p>Also, there are many other lifestyle modifications that he could possibly make that may decrease his risk for the development of diabetes such as physical activity, weight reduction, and dietary modification.</p>

Adapted from McMaster Evidence-based Clinical Practice Workshops

Diagnosis

Diagnosis Core Concepts

Notes on learning and teaching about **Diagnosis**:

1. **Background:** The importance of diagnostic reasoning cannot be overstated for teachers and users of the medical literature, yet this is an area in which many feel uncomfortable. The key concepts here can be separated into categories of 1) study design (prospective comparison to reference standard) for those wishing to review the original literature about diagnostic tests and 2) diagnostic reasoning and using likelihood ratios.
2. Assessing the validity of a study regarding a diagnostic test (risk of bias)
 - Optimal study design: prospective comparison to reference standard
 - Representative study group
 - Uniform comparison to a reference standard
 - Blinding of those interpreting the new test and reference standard (stumbling block: learners may confuse this concept with blinding in an RCT)
3. Understanding diagnostic test results
 - Sensitivity / Specificity
 - Likelihood Ratios
4. Diagnostic thinking (how do we use diagnostic test results)
 - Pretest probability and/or prevalence
 - Using a likelihood ratio (including use of the nomogram)
 - Posttest probability
 - Action threshold
 - Role of patient values and preferences as well as clinical circumstance in decision-making
5. Additional topics
 - Interobserver agreement / agreement beyond chance (Kappa)
 - Positive and negative predictive values and the downsides of using these test characteristics
 - Clinical Decision Rules
 - Understanding the role of randomized trials for studying implementation of a diagnostic test strategy (e.g. an intervention to use Ottawa ankle rule to decrease ankle x-rays in an emergency department)

Diagnostic Test Critical Appraisal Form

Citation:

How serious is the risk of bias?	
Did participating patients constitute a representative sample of those presenting with a diagnostic dilemma?	
Did investigators compare the test to an appropriate, independent reference standard?	
Were those interpreting the test and reference standard blind to the result of the other test ?	
Did all patients receive the same reference standard irrespective of the results of the test?	
What are the results?	
What are the likelihood ratios for the various possible test results?	
How can I apply the results to patient care?	
Will the reproducibility of the test results and its interpretation be satisfactory in my clinical setting?	
Are the study results applicable to patients in my practice?	
Will the test results change my management strategy?	
Will patients be better off as a result of the test?	

Adapted from McMaster Evidence-based Clinical Practice Workshops and Users' Guide to the Medical Literature 3rd Ed.

Differential Diagnosis Critical Appraisal Form

Citation:

How serious is the risk of bias?	
Did the study patients represent the full spectrum of those with this clinical problem?	
Was the diagnostic evaluation definitive?	
What are the Results?	
What were the diagnoses and their probabilities?	
How precise are the estimates of disease probability?	
How can I apply the results to patient care?	
Are the study patients and clinical setting similar to mine?	
Is it unlikely that the disease possibilities or probabilities have changed since this evidence was gathered?	

Adapted from McMaster Evidence-based Clinical Practice Workshops and Users' Guide to the Medical Literature 3rd Ed.

Screening Critical Appraisal Form

Citation:

How serious is the risk of bias?	
Is there randomized controlled trial evidence that the intervention benefits people with asymptomatic disease?	
What are the recommendations, and will they help you in caring for patients?	
Were the data identified, selected, and combined in an unbiased fashion?	
What are the benefits?	
What are the harms?	
How do benefits and harms compare in different people and with different screening strategies?	
What is the effect of individuals' values and preferences?	
What is the effect of uncertainty associated with the evidence?	
What is the cost-effectiveness?	

Adapted from McMaster Evidence-based Clinical Practice Workshops and Users' Guide to the Medical Literature 3rd Ed.

Diagnostic Test Worksheets

Definitions and the 2x2 table

		"Reference Standard" Result			
		Condition Present	Condition Absent		
Test Result	Positive Test	True Positive (a)	False Positive (b)	TP + FP a+b	PPV* = $a/(a+b)$ Of patients who test positive, the proportion who have disease
	Negative Test	False Negative (c)	True Negative (d)	TN + FN c+d	NPV* = $d/(c+d)$ Of patients who test negative, the proportion without disease
		TP + FN a+c	TN + FP b+d	a+b+c+d	Prevalence = $(a+c)/(a+b+c+d)$
		Sensitivity = $a/(a+c)$ Of patients with disease, the proportion who test positive	Specificity = $d/(b+d)$ Of patient without disease, the proportion who test negative	Accuracy = $(a+d)/(a+b+c+d)$ Of all patients in the study population, the proportion with true test results	Of all patients in the study population, the proportion with disease

(*test is affected by disease prevalence)

PPV = Positive Predictive Value. NPV = Negative Predictive Value

Sensitivity

PID = Positive In Disease

In a test with high sensitivity, this test will find most patients who have the disease.

In a test low sensitivity, many patients will disease will be missed (many false negative results)

SnNOUT = a Sensitive test with a Negative test result, rules OUT disease

Specificity

NIH = Negative In Health

In a test with high specificity, this test will correctly label patients with the disease

In a test with low specificity, this test will incorrectly label patients as having the disease (many false positives)

SpPIN- a Specific test with a Positive test result rules IN disease

Mathematical demonstration of how prevalence affects Positive Predictive Value:

Population size = 100,000, Sensitivity =90%, Specificity =90%

disease prevalence = 1%				disease prevalence = 0.1%			
	(+)	(-)		(+)	(-)		
	disease	disease		disease	disease		
(+) test	900	9,990	PPV= 8.3%	90	9,990	PPV= 0.9%	
(-) test	100	89,100		10	89,910		
	1,000	99,000	100,000	100	99,900	100,000	
11 false (+) for every true (+)				111 false (+) for every true (+)			

Likelihood Ratios:

Combine sensitivity and specificity into one measure

Conceptual Framework: How good is a diagnostic test in discriminating between patients with disease and those without disease?

Definitions:

- a likelihood ratio is a ratio of likelihoods
- likelihood of a disease based on a specific test result
 - (e.g. negative rapid strep test, intermediate VQ scan, WBC > 20,000 for appendicitis)
- a likelihood ratio compares the likelihood of a particular test result in patients with disease to the likelihood of that same result in patients without that same disease
- a likelihood ratio of 10 means that this specific test result is ten times more likely to occur in patients with disease than patients without disease
- a likelihood ratio modifies your pretest probability to generate a new, posttest probability

Math:

- By convention, a likelihood ratio compares the frequency that a specific test result (e.g. rapid flu negative) in patients **WITH** disease divided by patients **WITHOUT** disease
- Likelihood ratio for a **positive** test

$$\frac{\text{Proportion of patients who test positive who have disease}}{\text{Proportion of patients who test positive who do not have disease}}$$

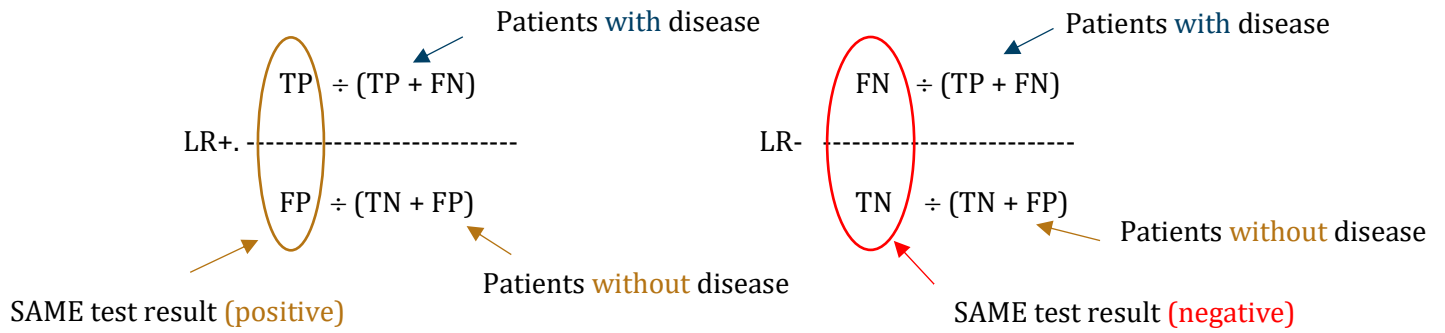
$$\frac{\text{Proportion of patients with positive rapid flu test who have influenza}}{\text{Proportion of patients with positive rapid flu test who do not have influenza}}$$

- Likelihood ratio for a **negative** test

$$\frac{\text{Proportion of patients who test negative who have disease}}{\text{Proportion of patients who test negative who do not have disease}}$$

$$\frac{\text{Proportion of patients with negative rapid flu test who have influenza}}{\text{Proportion of patients with negative rapid flue test who do not have influenza}}$$

- Going back to the 2x2 table



- Relating LR to Sensitivity and Specificity: sometimes you will want to convert sensitivity and specificity directly into LR (+) or LR (-) – this can be done with commonly used smartphone apps.

$$\text{positive likelihood ratio} = \text{sensitivity} / (1 - \text{specificity})$$

$$\text{negative likelihood ratio} = (1 - \text{sensitivity}) / \text{specificity}$$

Diagnosis Teaching Package Example

Learning Objectives:

1. To enhance skills in assessing the internal validity (risk of bias) for Diagnostic Test articles.
2. To learn to apply the results of a Diagnostic Test paper:
 - Estimating and understanding pretest probabilities
 - Calculating likelihood ratios directly from the data in the paper
 - Considering the importance of action thresholds in your clinical application of a diagnostic test result
 - Deciding whether the test will help you treat, understand, and/or advise your patient

Problem-Based Educational Package:

1. Read the attached clinical scenario.
2. Define the clinical question and outline a strategy to find the evidence
3. Read the attached JAMA articles about diagnostic tests.
4. Critically appraise the attached paper
5. Decide whether the diagnostic test will help you in your decision-making for your patient.

Enclosed Materials:

1. Wells PS, Hirsch, J et al. Accuracy of clinical assessment of deep-vein thrombosis. [Lancet 1995;345: 1326-30.](#)
2. Jaeschke R, Guyatt G, Sackett DL. Users' guides to the medical literature. III. How to use an article about a diagnostic test. A. Are the results of the study valid? Evidence-Based Medicine Working Group. [JAMA. 1994;271\(5\):389-391. doi:10.1001/jama.271.5.389](#)
3. Wilson MC, Hayward RS, Tunis SR, Bass EB, Guyatt G. Users' guides to the Medical Literature. VIII. How to use clinical practice guidelines. B. what are the recommendations and will they help you in caring for your patients? The Evidence-Based Medicine Working Group. [JAMA. 1995;274\(20\):1630-1632. doi:10.1001/jama.274.20.1630](#)
4. Critical Appraisal Worksheet
5. Worksheets for calculating likelihood ratios
6. Diagnostic test summary sheets

Worksheets for Calculating Likelihood Ratios:

Context:

In many instances, the papers you find will not describe the test properties in terms of likelihood ratios. You will need to calculate them or convert them yourself from the data given. The following pages are to help you do this. **These pages (Worksheets for Calculating Likelihood Ratios) are optional, however, they have helped some learners in the past.**

Contents:

- [Worksheet I: Calculating the Likelihood Ratios for the Clinical Assessment](#)
- [Worksheet II: Calculating the Likelihood Ratios for Ultrasound](#)
- [Formulas and Definitions Diagnostic Testing: Summary Sheet](#)

Clinical Scenario (A very real story)

It is your first experience as a general medicine ward attending. It is Saturday night and your team calls you because they are having trouble getting an ultrasound for one of the new admissions. The story unfolds as follows:

The patient is a 69-year-old male who has coronary artery disease, diabetes mellitus and is s/p a remote stroke several years ago. His baseline function is that he can transfer from bed to wheelchair, but he cannot ambulate on his own. His wife brought him to the ED this evening because he was “just not right,” but she denies any particular precipitating events. In the ED, he is noted to be pleasant but unable to offer history. His physical exam is notable for left hemiplegia which is consistent with prior reports in his medical record from his primary care physician. He also is anemic (HCT 30; MCV 78). In the ED, it is reported that he has left leg swelling, and he is admitted for “presumed DVT.” Your resident calls radiology to arrange an ultrasound this evening to help establish the diagnosis. The radiologist refuses, remarking that “it won’t change your management tonight anyway.”

Being a July attending, you come in to examine the patient with your team. Together, you measure his thigh and calf circumference. The left thigh is 2cm larger than the right in both measurements. He has no erythema but has bilateral peripheral pitting edema. He moans when you examine his left leg. However, there is no focal tenderness and no cords. He is guaiac positive with brown stool on digital rectal exam. His Hct 2 months ago was 40.

Your team is furious that the radiologist won’t come in to do the ultrasound. By now, it is 11 p.m. on a Saturday night. You phone the on-call attending radiologist. As it happens, he is the chief of radiology, and he tells you that he agrees that there is no reason to do the ultrasound tonight. You state that you are uncertain as to whether there is a clot. Furthermore, the patient appears to have rectal bleeding, and you are concerned about using anticoagulation in this setting unless it is absolutely necessary. You would like to be more certain of the diagnosis to make the safest decision. He holds his line and states that someone will be in tomorrow. In a final blow to your ego, he states “Call the chief of medicine and wake him up,” if you are not happy with the answer. You look up the pager number of your chief and wonder what to do...

Prior to reading the paper, please answer the following:

1. What is the pretest probability of DVT in this patient? (Write down your estimated percentage and bring it with you to the teaching session!)
2. Consider what questions you typically ask, what information you look for in a patient’s health record, and what physical examination items you perform. How good are history and physical examination findings at predicting the presence of a lower extremity DVT (i.e. how good are they as diagnostic tests?)
3. How much do you think the ultrasound will change the likelihood of a DVT in this patient if the result is positive? What if the ultrasound is negative?
4. How sure will you have to be that he has a DVT before you will be comfortable ordering anticoagulation? (or maybe you wouldn’t anticoagulate him at all?) (Record this as a percentage – how high does the probability/likelihood or DVT need to be?)
5. What arguments can you make in favor of the 11 p.m. ultrasound? What arguments might be against the 11 p.m. ultrasound?

Worksheet I:

Calculating the Likelihood Ratio for the Clinical Model for Evaluation of DVT

Table 2 Page 1328

First: you have to complete the 3x2 table below:

(Numbers are taken from table 2 by adding up the numbers of outcomes and patients in each group)

Clinical Model	DVT Present by Venogram	DVT absent by Venogram
High prob		
Moderate prob		
Low prob		
Totals	135	394

You can determine the likelihood ratio for a high probability assessment by just asking two questions:

1. How likely is a high probability assessment in people who really have DVT? (proportion 1)
2. How likely is a high probability assessment in people who DON'T have DVT? (proportion 2)

The Likelihood Ratio is simply a comparison of proportion 1 (disease) to proportion 2 (no disease). high probability assessment in disease/ high probability assessment without disease (again, notice – likelihood ratio is based on a single test result – in this case, high probability clinical assessment).

Next, from each possible test result (High, Moderate, or Low prob) calculate the Likelihood Ratio below

Clinical Model	Proportion 1 (patients with DVT)	Proportion 2 (patients without DVT)	Likelihood Ratio Proportion1/ proportion 2
High prob	__/135 =	__/394=	__/__=
Moderate prob	__/135=	__/394=	__/__=
Low prob	__/135=	__/394=	__/__=

Worksheet II:

Calculating the Likelihood Ratio for Ultrasound for Evaluation of DVT

For the following exercise, we will use Table 3 pg 1328 for Three Groups Combined

Set up the 2x2 table for **Proximal DVT**.

Combined Groups: Proximal DVT		
	(+) DVT	(-) DVT
(+) US		
(-) US		
	103	353

For Three Groups Combined, calculate the likelihood ratio for Ultrasound for **Proximal DVT**:

Ultrasound Result	Proportion 1 (DVT present)	Proportion 2 (DVT absent)	Likelihood Ratio (Proportion 1 / Proportion 2)
(+) US	___ / 103 =	___ / 353 =	LR(+) = ___ / ___ =
(-) US	___ / 103 =	___ / 353 =	LR (-) = ___ / ___ =

Set up the 2x2 table for **All DVT**.

Combined Groups: ALL DVT		
	(+) DVT	(-) DVT
(+) US		
(-) US		
	121	353

For Three Groups Combined, calculate the likelihood ratio for Ultrasound for **All DVT**:

Ultrasound Result	Proportion 1 (DVT present)	Proportion 2 (DVT absent)	Likelihood Ratio (Proportion 1 / Proportion 2)
(+) US	___ / 121 =	___ / 353 =	LR(+) = ___ / ___ =
(-) US	___ / 121 =	___ / 353 =	LR (-) = ___ / ___ =

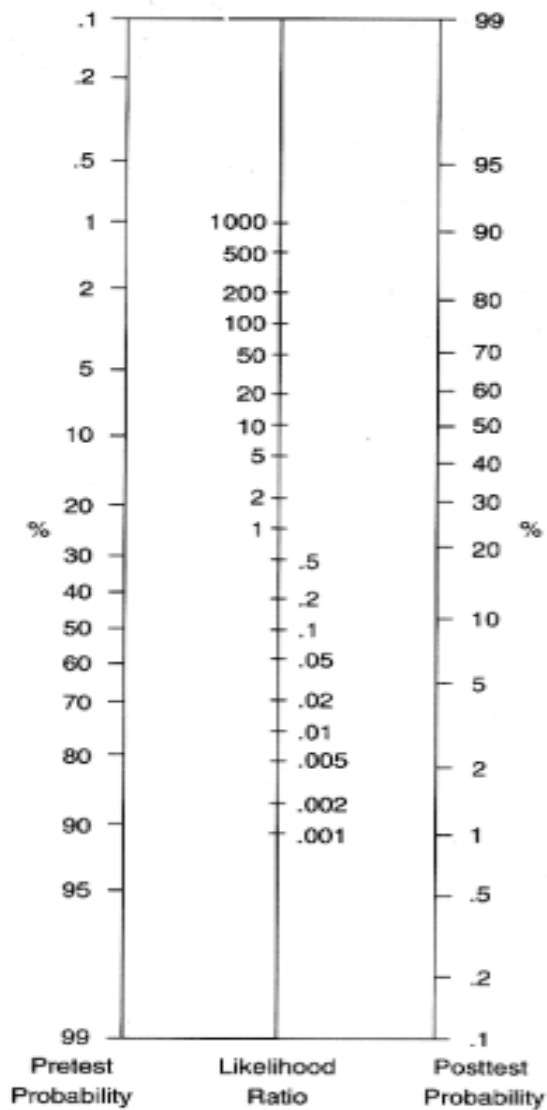
Application of Likelihood Ratios:

Write down the pretest probability of DVT: _____ (teacher will help with this)

What is the likelihood ratio of the test result you will use: _____(teacher to assign)

Draw a straight line from the pretest probability (left-most vertical line) through the likelihood ratio number (middle vertical line) to the posttest probability (right-most vertical line).

Fagan's Nomogram: tool to convert pretest to posttest probabilities by using a likelihood ratio



Adapted from Fagan TJ. Nomogram for Bayes's theorem [N Engl J Med Jul 31, 1975; 293\(5\):257](#). From Jaeschke, R. Users' guide to the medical literature: III. How to use an article about diagnostic test: B. What are the results and will they help me in caring for my patients? [JAMA Mar 2, 1994; 271\(9\):703-7](#) [Permission granted to reproduce image.]

What is the new, posttest probability of DVT in your patient: _____

What is your next step in managing this patient:

Diagnosis Teaching Package ANSWERS and Teaching Tips

Clinical Question:

What is the accuracy of clinical assessment and ultrasound for the diagnosis of suspected deep vein thrombosis (DVT)?

Population: Patients with suspected DVT

Intervention: clinical exam and also ultrasound

Comparison: The Reference Standard for diagnosis of DVT: contrast venography

Type of study you would prefer to find: a prospective trial in which there is a blind comparison of the tests we care about (clinical exam and ultrasound) to the reference standard for diagnosing DVT (contrast venography)

MEDLINE via PubMed Search

Access PubMed through the link at your institution. Or, you can find the freely accessible version of PubMed here <https://pubmed.ncbi.nlm.nih.gov/>, though please note that you may lose accessibility of full-text articles.

There are many ways to search PubMed. Let's try using the Clinical Queries, accessible from PubMed's homepage.

Enter: Deep Vein Thrombosis AND ultrasound

Switch the Clinical Study Category from Therapy to Diagnosis, and select Narrow for the scope.

PubMed will display the first 5 citations based on a best match relevancy algorithm. At the bottom of that list, you can click to view the full set of results in PubMed. Once you click on those results, you can

further limit to by language, year, or other limits.

deep vein thrombosis AND ultrasound

COVID-19 Articles

Category: Treatment

5 of 135 results

Diagnosis and resection of a giant ovarian cyst presenting in a young patient with contralateral back pain and lower limb deep vein thrombosis.
Perera D, et al. BMJ Case Rep. 2021. PMID: 33727290

Pathogeny of cerebral venous thrombosis in SARS-Cov-2 infection: Case reports.
Guendouz C, et al. Medicine (Baltimore). 2021. PMID: 33725828

SARS-CoV-2 and finding of vein thrombosis: can IMPROVE and IMPROVEDD scores predict COVID-19 outcomes?
Greco S, et al. Eur Rev Med Pharmacol Sci. 2021. PMID: 33660832
Free article.

Severe acute respiratory syndrome coronavirus 2-induced acute aortic occlusion: a case report.
Minalyan A, et al. J Med Case Rep. 2021. PMID: 33653414
Free PMC article.

Incidence of thrombotic complications and overall survival in hospitalized patients with COVID-19 in the second and first wave.
Dutch COVID & Thrombosis Coalition, et al. Thromb Res. 2021. PMID: 33535120
Free PMC article.

See all (135)

Clinical Study Categories

Category: Diagnosis

Scope: Narrow

5 of 969 results

Cost-effectiveness of magnetic resonance imaging for diagnosing recurrent ipsilateral deep vein thrombosis.
van Dam LF, et al. Blood Adv. 2021. PMID: 33661297
Free PMC article.

Diagnostic accuracy of four different D-dimer assays: A post-hoc analysis of the YEARS study.
Hamer HM, et al. Thromb Res. 2021. PMID: 33626463
Free article.

Assessing the thrombosis risk of peripherally inserted central catheters in cancer patients using Caprini risk assessment model: a prospective cohort study.
Feng Y, et al. Support Care Cancer. 2021. PMID: 33594508

Incidence and predictors of occult preoperative deep vein thrombosis at radical cystectomy for urothelial carcinoma.
Miest TS, et al. Can Urol Assoc J. 2021. PMID: 33591896

COVID-19 Infection in Critically Ill Patients Carries a High Risk of Venous Thrombo-embolism.
Bellmunt-Montoya S, et al. Eur J Vasc Endovasc Surg. 2020. PMID: 33583710
Free PMC article.

See all (969)

This column displays citations filtered to a specific clinical study category and scope. These search filters were developed by Haynes RB et al. See more filter information.

This column displays citations about the 2019 novel coronavirus filtered by

←Your search will look something like this.

You scan the abstracts, selecting those that are about diagnosis of DVT using clinical exam and ultrasound. You pick the article by Wells et al. because it seems to most directly answer your question and to be methodologically valid (although it is sometimes hard to tell by the abstract).

Discussion of Pretest Probabilities:

You cannot discuss a diagnostic test without discussing pretest probability. You might engage your learners by having them “guesstimate” the probability of disease in the case presented and then compare the various assessments. This is a common stumbling block as learners are often uncomfortable with estimating the probabilities. Learners will want to know where these numbers “come from.” In general, the pre-test probability is a combination of clinical expertise and knowledge of the literature/epidemiology of the disease in your practice setting. As such, novice learners will struggle the most. It’s OK that pre-test probability estimates will vary among your learners – this is what happens in the real world; two different clinicians may see the same patient with shortness of breath and derive different probabilities of heart failure vs. COPD exacerbation vs. PNA. The pre-test probabilities each clinician develops for a certain disease affects subsequent decisions. For example, if physician A feels patient Y has a 5% probability of a pneumonia, her next steps are likely to be much different than physician B who feels patient Y has a 25% probability of pneumonia. The same will hold true with the application of likelihood ratios.

Tips:

- Have learners write down his or her pre-test probability on a small piece of paper (secret ballot) before sharing their “answer” with the larger group. This will minimize the “herd effect” of everyone guessing the same value. Set aside these estimates to use later after you have calculated the likelihood ratios.
- You may also want to vary the scenario to create a picture that is very different (much more or less likely to have the disease in question). By doing this, you can most effectively show the varied impact likelihood ratios on altering the probability of disease (i.e. the magnitude of the change between pre- and post-test probabilities).
Or, if learners had reported varied pre-test probabilities, you can use several to demonstrate this effect. Again, remind students it’s OK if pre-test probabilities vary among them – this is real world!

The Critical Appraisal Exercise Answers:

Citation: Accuracy of clinical assessment of deep-vein thrombosis ([Lancet, 1995; 345: 1326-30](#))

Clinical Question: What is the accuracy of clinical assessment and ultrasound for the diagnosis of suspected deep-vein thrombosis (DVT)?

How serious is the risk of bias?	
<u>Did participating patients constitute a representative sample of those presenting with a diagnostic dilemma?</u>	<p>Possibly.</p> <p>The patients are outpatients referred to three centers for evaluation of suspected DVT. Hospitals were university-affiliated centers in Canada and Italy.</p> <p><u>Exclusion criteria:</u> previous DVT or PE, suspected PE at presentation, inability to take contrast, anticoagulation for >48h, below the knee amputation and pregnancy.</p> <p><u>Inclusion criteria:</u> outpatients with suspected DVT for <60 days</p> <p>Because this is a population referred to a set of university associated medical centers, there may be referral bias, diagnostic workup bias and/or diagnostic suspicious bias. All of these biases will favor a population more enriched with the disease and, perhaps more likely to have a positive ultrasound. (The end result of having more people who will test positive → overestimate of sensitivity and underestimate of specificity.)</p>
<u>Did investigators compare the test to an appropriate, independent reference standard?</u>	<p>Yes.</p> <p>All patients underwent a 3 step process:</p> <ol style="list-style-type: none"> 1. Clinical model applied to estimate pre-test probability. 2. Ultrasound 3. Venography on the same day, if possible. If not, it was done the following day (Reference Standard)
<u>Were those interpreting the test and reference standard blind to the result of the other test?</u>	<p>Yes.</p> <p>US and venogram reports were interpreted by a panel of 3 or more observers blinded to the diagnostic test results or patient history.</p> <p>Discussion and consensus resolved disagreements.</p>
<u>Did all patients receive the same reference standard irrespective of the result of the test?</u>	<p>Yes.</p> <p>Venograms were done on all subjects, regardless of any clinical history, physical findings or ultrasound reports.</p>
What are the results?	
What are the results?	<p><u>Overall Prevalence of DVT in all study patients</u></p> <p>Of 529 patients, 135 had DVT (26%)</p> <p>Of these 113 had proximal DVT (21%)</p> <p><u>Prevalence of DVT according to subgroups determined by the clinical model:</u></p> <p>High pre-test probability 85%</p> <p>Moderate pre-test probability 33%</p> <p>Low pre-test probability 5%</p>
What were the Likelihood ratios?	<p>(See Worksheets attached)</p> <p><u>Clinical Exam:</u> We can use clinical criteria as a diagnostic test to stratify patients into categories of high (LR 16.2), moderate (LR 1.5) and low probability (LR 0.17). For</p>

	<p>patients in the moderate probability category, we gain no valuable diagnostic information and we must go on to a second test (e.g. venography).</p> <p><u>Ultrasound</u>: In many patients, an ultrasound will greatly modify our pre-test probabilities. A negative US will essentially rule out DVT (LR- 0.22) whereas if it is positive it is essentially diagnostic for DVT (LR+ 39). The properties of the test are even better for evaluation of proximal DVT (LR+ 44.5; LR- 0.11)</p>
How can I apply the results to patient care?	
Will the reproducibility of the test result and its interpretation be satisfactory in my setting?	<p>Probably yes.</p> <p>Both Doppler ultrasonography and venography are tests that require consistency and skill in both performance and interpretation. However, it does seem that the criteria applied for both abnormal ultrasound (lack of full compressibility) and abnormal venogram (constant luminal filling defect in at least 2 projections) are appropriate and accepted standards.</p> <p>It is likely that these results could be reproduced at university-affiliated centers, such as those involved in the study. However, possibly at smaller centers, there may be less experience with these procedures and the results and interpretation may be more variable.</p>
Are the results applicable to my patient?	<p>Yes. My patient presented to the ED of a university-affiliated hospital and was sick enough to be admitted to the hospital. Unfortunately, there is no description of patient demographics to define the racial, gender or socioeconomic make-up of the group studied. We do know that the centers spanned 2 countries (Canada and Italy) and that the DVT frequencies varied from institution to institution. The implication is that there may be important differences in diverse populations of patients.</p>
Will the test results change my management strategy?	<p>Yes, definitely.</p> <p>The clinician will certainly take different actions depending on whether the diagnosis of DVT is confirmed, ruled out or remains in doubt. Specifically, the clinician will need to decide whether or not to anticoagulate the patient.</p> <p>When clinical exam and ultrasound <u>agree</u> (high pretest probability and + US or low pretest probability and –US), we will effectively rule in or rule out disease, respectively. Clinicians can feel comfortable making decisions about anticoagulation on this basis alone.</p> <p>When clinical examination and ultrasound are in <u>disagreement</u> (high pre-test probability and –US or low pre-test probability and +US), additional testing such as venogram might be necessary.</p>
Will patients be better off as a result of the test?	<p>Yes, definitely. The combination of clinical exam and ultrasound provides a safe, non-interventional way of confirming or excluding DVT. This is essential in the treatment of disease and in the prevention of migration of the clot (e.g. to prevent pulmonary embolism). This combination of tests is far safer than the reference standard, venogram.</p>

Adapted from McMaster Evidence-based Clinical Practice Workshops

ANSWERS Worksheet I: Calculating the LR for the Clinical Model for evaluation of DVT

(Numbers are taken from table 2 by adding up the numbers of outcomes and patients in each group)

Clinical Model	DVT Present by Venogram	Proportion 1 (DVT present)	DVT absent by Venogram	Proportion 2 (DVT absent)	Likelihood Ratio (Proportion1/ proportion 2)
High prob	72	$72/135 = 0.533$	13	$13/394 = 0.33$	$0.53/0.033=16.2$
Moderate prob	47	$47 /135 = 0.35$	96	$96/394 = 0.24$	$0.35/0.24=1.5$
Low prob	16	$16 /135 = 0.12$	285	$285/394 = 0.72$	$0.12/0.72=0.17$
	135		394		

ANSWERS Worksheet II: Calculating the LR for the US for evaluation of DVT

Using numbers from table 3 on page 1328, please complete the following 2x2 table:

Combined Groups: Proximal DVT			Combined Groups: ALL DVT			
	(+) DVT	(-) DVT		(+) DVT	(-) DVT	
(+) US	92	7	(+) US	94	7	
(-) US	11	346	(-) US	27	346	
	103	353	456	121	353	474

For Three Groups Combined calculate the likelihood ratio for determining **Proximal DVT** using Ultrasound:

Ultrasound Result	(+) DVT present	Proportion 1 (DVT present)	(-) NO DVT	Proportion 2 (DVT absent)	Likelihood Ratio (Proportion 1 / Proportion 2)
(+) US	92	$92 /103 = 0.89$	7	$7 /353 = 0.020$	$LR(+)$ $0.89 / 0.020 = 44.5$
(-) US	11	$11 /103 = 0.11$	346	$346 /353 = 0.98$	$LR (-) = 0.11/0.98 = 0.11$

Ultrasound Result	(+)DVT	Proportion 1 (DVT present)	(-)NO DVT	Proportion 2 (DVT absent)	Likelihood Ratio (Proportion 1 / Proportion 2)
(+) US	94	$94 /121 = 0.78$	7	$7 /353 = 0.020$	$LR(+)$ $0.78/0.020 = 39$
(-) US	27	$27 /121 = 0.22$	346	$346 /353 = 0.98$	$LR (-) 0.22/0.98 = 0.22$

For Three Groups Combined calculate the likelihood ratio for determining **All DVT** using Ultrasound:

What is the new, posttest probability of DVT in your patient: _____

What is your next step in managing this patient:

Once the learner determines the new (posttest) probability of DVT in this patient, ask the learner what their next step will be in the care of this patient.. This is the whole point of the exercise – demonstrating the utility of diagnostic tests and ,specifically, how likelihood ratios impact in-the-moment clinical decision-making.

Important questions to ask:

1. Did the diagnostic test (e.g. clinical model or ultrasound) change your pretest to posttest probability significantly?
2. Is your new posttest probability above your treatment threshold or below your observation threshold? Or does the probability of DVT remain in the zone of uncertainty?

Resolution of the Clinical Scenario:

The team decided that we would not be willing to anticoagulate this patient without greater certainty as to the diagnosis of DVT. The ultrasound result, whether positive or negative, will substantially alter the patient's probability of having a DVT. In addition, we considered the other possible option of placing a greenfield filter to prevent PE, if the patient's hematocrit dropped further and we could not anticoagulate him. We got an interventional vascular radiologist to agree that she would place the filter if the US revealed clot and if the patient could not be anticoagulated. Armed with this information, we made a final call to the on-call chief of radiology. He agreed to send someone in to perform the ultrasound that evening. The ultrasound was negative, ruling out a DVT, thus sparing the patient risky anticoagulation.

Tips:

1. Sessions often work better if teacher demonstrates use of the nomogram first. Start by choosing a pretest probability. You could use your own pretest probability for this case, one of your learner "guestimates," or an estimated average of all of your learners pretest probabilities. Then choose one of the likelihood ratios everyone calculated. We suggest using Moderate Probability Clinical Model. The patient in the scenario falls into the Moderate Probability Category from Table 1 on p 1328. He has 2 Major points (paresis and swollen calf and thigh) but no minor points. Clinical Examination for patients in this category is not a useful diagnostic test (LR=1.5).
1. Use the nomogram to determine the posttest probability. Then ask the group, what should we do next?
2. For ease of discussion, it is often best to pick one pretest probability for the entire group to use. Oftentimes, the estimated average of the learners pretest probability works well.
3. Divide the group into smaller groups and assign them specific test results (e.g. ultrasound + for Proximal DVT to group 1, ultrasound negative for Proximal DVT to group 2, ultrasound + for ALL DVT to group 3, ultrasound negative for ALL DVT to group 4). Choose as many groups as you feel is best and you have time for. When you assign the specific test result instruct the group to determine the posttest probability and what

the next step in their management of the patient will be (another test, start anticoagulation, no anticoagulation).

4. If there is more time, you can change the clinical scenario substantially such that the pretest probability for DVT will be much higher and much lower than the case presented. Again, assign specific test results to learners so that they may determine the posttest probability and decide on next steps. In this way, you can demonstrate how various pretest probabilities affect the ultimate posttest probability and next patient management actions.
5. Calculating the LR for US in the diagnosis of suspected DVT shows you that US is a very good test (LR (+) \geq 40 and LR(-) \leq 0.2 for proximal DVT) and it will help us make our diagnosis. However, the radiologist in the scenario forces you to address the next question. What will you do with the information when you get it?
6. Be very careful about timing when teaching diagnostic testing. It takes learners a long time to go through the transformation of the data into likelihood ratios. You may need to only quickly review the validity criteria so that you can focus your energy on the case, the results section and how you will use likelihood ratios to make the clinical decisions.
7. It is very effective to always begin with a 3x2 table such as the Clinical model which has 3 possible results (High, moderate and low probability). This forces the teaching point that you cannot calculate sensitivity and specificity when you have more than 2 possible results of your test.

ANSWERS: Extra Practice- Likelihood Ratios from Table 3:

You can use the Sensitivities and Specificities in Table 3 on page 1328 to derive the following likelihood ratios for extra practice. For those of you who like formulas, you can convert directly to LR(+) and LR(-) using:

$$\text{positive likelihood ratio} = \text{sensitivity} / (1 - \text{specificity})$$

$$\text{negative likelihood ratio} = (1 - \text{sensitivity}) / \text{specificity}$$

Pre-Test Probability (Derived by Clinical Model)	Likelihood Ratios for Proximal DVT		Likelihood Ratios for All DVT	
	LR(+)	LR(-)	LR (+)	LR (-)
High Risk	So very high*	0.06	So very high*	0.09
Moderate Risk	83	.17	61	.39
Low Risk	40	0.2	33.5	0.34
All Groups Combined	44.5	0.11	39	0.22

**because the specificity was 100% (i.e. there no false + tests), you can only calculate the LR+ if you consider the confidence interval around zero and use a number other than 0.*

Warnings:

1. You can use these formulas only if you are looking at a dichotomous test (Positive or Negative).
2. From a teaching perspective, it does not show your learners what a likelihood ratio really means- it is simply a formula that they will likely forget when they need it!

Summary Methods SR MA Guidelines

Systematic Review Core Concepts

Notes on learning and teaching about **Systematic Review and Meta-analysis**:

1. **Background:** Systematic Reviews / Meta-analyses are examples of a summary methodology that summarizes original, individual trials. The subjects of a systematic review can be thought of as the papers that are brought together to answer a particularly focused question. Thus, the methods for a systematic review or meta-analysis will talk a lot about the papers that were collected for the study including how they were identified, selected, graded for quality, and possibly (in the case of meta-analysis) combined. These summaries can be on many different kinds of clinical questions (e.g. therapy, diagnosis, or prognosis).
2. Key concepts and terms:
 - a. Narrative Review vs. Systematic Review vs. Meta-analysis
 - b. Summary estimate or pooled estimate of effect
 - c. Forest Plots including point estimates, confidence intervals, and line of no difference
 - d. Heterogeneity including I^2 and yes/no tests for heterogeneity with p-values
 - e. Reporting bias including publication bias and funnel plots
 - f. GRADE (Grading of Recommendations, Assessment, Development, and Evaluation)
3. Additional Topics
 - a. Network Meta-analysis

Systematic Review

Citation:

Assessing the Credibility of the Systematic Review Process	
Did the review address a focused clinical question (i.e. can be framed in PICO format)?	
Was the search for relevant studies detailed and exhaustive?	
Were selection and assessment of studies reproducible?	
Was the risk of bias of the primary studies assessed?	
Did the review address possible explanations of between-study differences in results using prespecified hypotheses?	
Did the review describe a process to assess confidence in effect estimates? (e.g. GRADE tool to assess quality of the body of evidence)	
Understanding the Summary Estimate of a Meta-analysis	
What is the magnitude of treatment effect? (what is the pooled estimate?)	
How precise are the results? (i.e. confidence interval around the pooled estimate)	
Rating Confidence in the Estimates (the Quality of a Body of Evidence)	
How serious is the risk of bias in the body of evidence?	
Are the results consistent across studies? (i.e. heterogeneity or inconsistency)	
Do the results directly apply to my patient? (i.e. PICO, generalizability, indirectness)	
Is there a concern about reporting or publication bias?	
Are there reasons to increase or decrease the confidence of the rating? (<i>Randomized trials start high and observational studies start low</i>)	
Overall, what is the quality of the body of evidence by outcome? (High, moderate, low, very low)	
How can I apply the results to my patient care?	
Did the review present results that are ready for clinical application? (e.g. patient important outcomes, absolute benefit /risk)	
Are the study patients similar to my patient and are likely benefits worth potential harms/costs?	

Adapted from McMaster Evidence-based Clinical Practice Workshops and the Users' Guide to the Medical Literature 3rd Ed.

Comparing different types of reviews

Teaching Table for Comparing Several Common Kinds of Summary Articles:

Background: Efficient application of the medical literature requires that we make optimal use of articles in which the authors have combined information from multiple individual original studies. Systematic Review and Meta-analyses are increasingly available to pull together a comprehensive set of available original articles on a particular focused clinical question. Clinical Practice Guidelines also pull together multiple original sources; however, the starting point is a clinical problem made up of many individual, focused clinical questions. Clinical Practice Guidelines are frequently based upon the work of systematic reviews and meta-analyses.

Systematic Review, Meta-analysis and Clinical Practice Guidelines: The following table can be used in a teaching setting to help your learners understand the differences between these summary methodologies. One way to use the table is to begin with the headings on top and create the table interactively with the learners answering questions about the different types of studies.

	Unsystematic Review	Systematic Review	Meta-analysis	Clinical Practice Guideline
(also known as)	Narrative Review	Qualitative Review	Quantitative Review	
Evidence Summary?	Maybe (at the discretion of the author)	Yes	Yes	Yes
Is this review based on a focused clinical question?	No A narrative review is usually based on a clinical problem (e.g. Review of GI Bleeding)	Yes	Yes	No A family of questions related to a complex clinical problem (e.g. diabetes care)
Does this kind of review have a methods section?	No. Narrative reviews are written in the style suggested by the journal and the author without an explicit methodology.	Yes. Systematic Reviews have methods sections that include comment on the following core elements of article selection: How they found the evidence (comprehensive search strategy) How they determined the quality of the evidence (validity check)	Yes. Meta-analyses Have methods sections that include everything in a Systematic Review (comprehensive search strategy and validity check) AS WELL AS a Summary Statistic (combining the data from individual studies based on precision including sample size and variability)	Yes. A Clinical Practice Guideline allows an integration of how to approach a clinical problem. Methods include a description of comprehensiveness, quality, validity and also the process for making recommendations when there is no evidence
Who's viewpoint is represented?	The authors (Expert Model / Authority)	Evidence Model: this is simply a systematic collection and 'grading' of scientific data	Evidence Model: based on a systematic review with a combining of data from individual studies into a summary statistic	Evidence Model and Expert Model combined: This includes both evidence and expert opinion when evidence is not available.

A Shorter Version of the Teaching Table for Systematic Review and Meta-analysis:

	Narrative Review	Systematic Review	Meta-analysis
Kind of question	Topic (e.g. GI bleed)	Focused Question	→ (same as SR)
Methods	None	Comprehensive Search	→ (same as SR)
		Screen articles for validity	
Results	None	Summary of Evidence (can be qualitative)	Summary Stat
Who's perspective	Authority Model	Evidence Model	→ (same as SR)

Meta-analysis, Decision Analysis, and Economic Analysis: Another set of summary methodologies includes decision and economic analysis. You can use a similar strategy for teaching about these kinds of papers. First, you point out that there are three different types of measures that researchers might make (outcomes, values, and costs). Then you can elicit examples from

Outcomes	Values	Costs
Examples:	Examples of different viewpoints (i.e. who's values?)	Examples:
# admits	Patients	Drug costs
time to symptoms	Hospital / administrators	personnel
time to C/C	parents	admission \$\$
side effects	physicians	lost work /wages
mortality	insurance	lost school/ work
	society	equipment

How these come together in summary articles:

Meta-analysis / Overview: summary of the outcomes literature

Decision Analysis: takes all the outcomes and considers the weight of values (outcomes x values)

Economic Analysis: takes outcomes, values, and costs

(at times, does not include values as below)

Types of cost studies:

Cost-benefit analysis: all outcomes are in monetary units and no value assigned

Cost-effectiveness: monetary cost compared with a clinical unit of efficacy

Cost-utility analysis: monetary costs compared with outcomes measured in terms of social value
(e.g. Cost per QALY)

Systematic Review Class Worksheets

Clinical Case Scenario:

Your patient is a 68-year-old woman with a history of known CAD, s/p Inferior MI 7 years ago. Following her MI, she did extremely well. She has been on aspirin and b-blocker therapy. For the past 2 years, she had been walking 2 miles four times each week without pain until 3 weeks ago when she noted exertional pain when she goes for her walk. She presented to the ER this evening when she had 4/10 substernal pain while sitting down at dinner. She denied SOB, N/V or diaphoresis.

In the ER, she was treated with aspirin, oxygen and sublingual nitroglycerine with complete resolution of her pain. Her initial physical examination was significant for HR 100 bpm, BP 105/70, clear lungs and heart exam without murmurs or gallops. She had no signs of CHF. Her presenting EKG showed anterior ST depression which resolved post nitroglycerine.

She is admitted to the hospital.

Question #1:

The intern is writing the admission orders and asks you whether to treat this patient with Heparin at this time. Do you want to treat this patient with heparin?

Yes

No

Not Sure

Question #2:

The medical student adds that the patient was told at the time of her previous MI that she had some blood in her stools. She doesn't recall receiving any transfusions for this and doesn't recall what work up was done because "she had so many tests at that time". You don't have access to these medical records. At this time, she is guaiac negative.

The patient is concerned about bleeding and would like to know more about her risk of bleeding. Because you care so much about the learning experience of your medical student, you suggest that she look to the medical literature to find out the answers to these questions.

What specific numbers do you want her to bring back to the team to help put into perspective the potential risks/ benefits of heparin therapy in this patient?

Task 1

Construct clinical question(s) from this case.

Construct your question

Patient, Production or Problem	Intervention, Prognostic Factor Exposure	Comparison Intervention (if appropriate)	Outcome you would like to measure or achieve	Type of Question you are asking	Type of Study you would want to find
How would I describe a group of patient's similar to mine?	Which main intervention, exposure, prognostic factor am I considering?	What is the main alternative to compare with the intervention?	What can I hope to accomplish, measure, improve, affect?	How would I categorize this question?	What would be the best study design in order to answer this question?
Search Strategy					
Search Strategy					

What is the research question for the paper and how does it compare to the clinical question?

--

Table for the Review of Systematic Reviews / Meta-Analysis

Questions	
Are the Results Valid? <u>Clinical Question</u> (Did the Review address a focused clinical question?)	
<u>Search Strategies</u> (Were the criteria used to select studies for inclusion appropriate?)	
<u>Search Strategies</u> (Is it likely that important, relevant studies were missed?)	
<u>Validity of included studies</u> (Were the studies included appraised for validity?)	
<u>Assessment of included studies</u> (Were the assessments of the studies included reproducible?)	
<u>Similarities of included studies</u> (Were the results of studies included similar study to study?)	

<p>What are the Results? (What is the Strength of the Outcomes?) What were the Overall Results of the Review?</p>	<p>Calculate the NNT for:</p> <ol style="list-style-type: none"> 1. Calculations for Risk of Death/MI during therapy (Table 2) 2. Calculations for Risk of Death/MI during therapy If you take out Holdright et al (Table 2) 3. Calculations for Death/MI 2-12 weeks: 4. Calculations for NNH: Risk of "major bleeding" Paragraph on page 813
<p>How precise are the results?</p>	
<p>Will the Results help me in caring for my patients? Are the results applicable to my patient?</p>	
<p>Were all clinically important outcomes considered?</p>	
<p>Are the benefits worth the potential harms and costs?</p>	

Systematic Review Example of Teaching Package **ANSWERS** (Therapy Unstable Angina)

Critical Appraisal Worksheet for Meta-analysis

Using Systematic Reviews / Meta-Analysis in clinical practice

What is the research question for the paper and how does it compare to the clinical question?

Clinical question and Research question are the same: In patients with Unstable Angina, is heparin plus aspirin more effective than aspirin alone in preventing MI and death?

We had the additional question: In patients with Unstable Angina, is heparin plus aspirin more likely than aspirin alone to cause bleeding?

Table for the Review of Systematic Reviews / Meta-Analysis

Questions	Adding Heparin to Aspirin reduces the incidence of myocardial infarction and death in patients with unstable angina (<i>JAMA</i> . 1996; 276: 811-815)
Are the Results Valid? Clinical Question (Did the Review address a focused clinical question?)	Yes —clinical question In patients with Unstable Angina, is heparin plus aspirin more effective than aspirin alone in preventing MI and death?
Search Strategies (Were the criteria used to select studies for inclusion appropriate?)	Search Strategy: keywords: heparin, aspirin, unstable angina Included articles in English and non-English articles. A manual search was done as well. Consultation with experts was also done. Inclusion criteria for studies: <ul style="list-style-type: none"> ➤ Design: random allocation ➤ Patient population: Unstable Angina or non-Q-wave MI ➤ Intervention: IV heparin + aspirin vs aspirin alone ➤ Outcomes: MI or mortality
Search Strategies (Is it likely that important, relevant studies were missed?)	Medline, reference lists, expert contacts were all used. Should have searched more than one database. ? whether they sought unpublished data. ? whether they searched conference abstracts
Validity of included studies (Were the studies included appraised for validity?)	This is a question of Therapy! Therefore, we should want to know: Was the assignment of patients random? (Yes) Was the randomization blinded? (Two studies were double-blind, two were single blind, two were not blinded at all) Were all of the patients who entered the trial properly accounted for and attributed at its conclusion? (“follow up is excellent” only one study reports 2 patients lost to follow-up at 12 weeks) Were there important co-interventions that might affect the outcome? (We don’t know. For example, there is no mention of whether b-blocker use was accounted for in each of the studies)
Assessment of included studies (Were the assessments of the studies included reproducible?)	Data abstraction and methodologic quality assessment were conducted in duplicate; one observer was blinded to journal, year of publication, authors, and institution. Disagreements were determined by consensus but no Kappas were reported.

<p>Similarities of included studies (Were the results of studies included similar study to study?)</p>	<p>Patients: Clearly defined criteria for the USA. All patients had either:</p> <ul style="list-style-type: none"> ➤ recent onset of prolonged or recurrent chest pain suggestive of ischemia (4 trials) ➤ increasingly severe pain with activity or rest pain (6 trials) ➤ pain within 48 hours of admission (5 trials) <p>Interventions: Varied across studies:</p> <ul style="list-style-type: none"> ➤ ASA dose range: 75 mg q od to 325 mg bid ➤ Heparin tx duration range was from 2→7 days ➤ PTT target was 1.5-2.0x normal <p>Results:</p> <ul style="list-style-type: none"> ➤ The collection of studies passes the eyeball test (intraocular view of Figure 1 shows favorable trends in each study) ➤ The random effects model was used for statistical pooling. This incorporates between study as well as within study variation. This is a more conservative estimate of the treatment effect than an analysis using the fixed effects model. It assumes that the patients are drawn from the “universe of heart disease”) rather than a very homogenous group. ➤ The formal test for homogeneity was done. They report a formal test for homogeneity for each outcome; chi-sq was not significant for each outcome except ischemic pain which did have a significant test for homogeneity. (Note:- TESTS FOR HOMOGENEITY ARE NOT POWERFUL STATISTICALLY! Most of these studies are not powered to show small differences.) 		
<p>What are the Results? (What is the Strength of the Outcomes?)</p>	<p>6 trials included: no statistical heterogeneity: Note: one of the trials (Holdright has a much higher MI/Mortality rate)</p>		
<p>What were the Overall Results of the Review? How precise are the results?</p>	<p>Outcome</p>	<p>Absolute risk Hep+asp (asp alone)</p>	<p>Relative Risk</p>
	<p>Death/MI during tx Death/MI (2-12 wk) Recurrent ischemia Revascularization Major bleeding</p>	<p>55/698 (68/655) 3/655 (10/698)</p>	<p>0.67 (0.44-1.02) 0.82 (0.56-1.20) 0.68 (0.4-1.17) 1.03 (0.74-1.43) 1.89 (0.52-7.65)</p>
	<p>Calculations for Death/MI during therapy Risk of death/MI: Absolute risk reduction = (68/655)- (55/968) = 0.104 – 0.079 = .025 NNT= 1/0.025=40 patients</p> <p>If you take out Holdright et al: Absolute risk reduction = (28/524)- (13/544)= .053-.024 = .0295 NNT= 1/0.0295 = 34 patients</p> <p>Calculations for Death/MI 2-12 weeks: Assumption from Table 2: risk of in hospital death approx 5% For patients treated only with Aspirin therapy (0.05) – (0.05* .82)= 0.05-0.41=0.009 1/0.009 = 111</p> <p>Calculations for Risk of “major bleeding” Absolute risk increase=</p>		

	$(10/698) - (3/655) = 0.009746$ $NNT = 1 / 0.0097 = 103$
Will the Results help me in caring for my patients? Are the results applicable to my patient?	<p>The studies included here are likely to be representative of the evidence that is available in 1998. The populations, interventions, outcomes are reasonably well described. Individual readers need to decide whether this is generalizable to their practice.</p>
Were all clinically important outcomes considered?	<p>Death and MI are aggregated- ideally, we would like to see them separately. They are measured for short (duration of therapy) and longer term (2-12 weeks after). Morbidity including bleeding, recurrent pain, and revascularization were looked at.</p>
Are the benefits worth the potential harms and costs?	<p>Harm: We have reviewed bleeding and revascularization issues in the results section. We need to treat between 34-40 patients to prevent 1 MI/Death and we will cause one "major bleed" for every 100 patients we treat</p> <p>Long-term studies are really needed to get a more precise estimate of benefit post hospitalization, adverse effects of major and minor bleeding.</p> <p>Alternative therapies: The effect of LMW heparin</p> <p>Cost: A full economic analysis has not been conducted</p>

Qualitative and Other Methods

Other Topics Core Concepts

Notes about learning and teaching about **Other Topics**:

1. In addition to the CORE areas, there are many iterations of critical appraisal exercises. In this section, we have given you some examples of commonly related topics that come up. Specifically, we have given you several examples of the following kinds of topics: Qualitative Research, Prevention, and Screening. These might be appropriate topics for more advanced learners or for taking home to try sometime in the future. These topics build on the knowledge and skills that are developed in the CORE areas but are slight deviations from the CORE.
2. Prevention. Prevention is frequently studied in a manner similar to therapy (i.e. with RCT methodology) because it is an intervention (e.g. aspirin to prevent MI). However, one of the main differences pertains to the relative balance of potential benefit to potential harms because prevention deals with individuals who are without the target disorder (primary prevention) or who are trying to prevent recurrence (secondary prevention) whereas a therapy is required to treat a present disorder to prevent related adverse outcomes.
3. Screening. Screening is also frequently studied by RCT because it too is an intervention. In this case, a diagnostic test is used as an intervention to screen for and identify early disease and ultimately to prevent adverse outcomes once the disorder is identified. However, screening can also be studied in terms similar to the study of a diagnostic test, if the question pertains to the ability of the diagnostic test to pick up the target disorder. In this case, then the methodology would more parallel a prospective comparison to a reference standard as in a classic diagnostic test study.
4. Qualitative methods. Qualitative research uses open-ended methodology (e.g. focus groups or in-depth interviews) to generate hypotheses and expand our thinking in the area of inquiry. Because we are more familiar with quantitative methods, teaching can be both challenging and fun. This section includes an example of several types of teaching exercises to offer some thoughts about how one might teach qualitative methodology.

Quantitative vs. Qualitative Methods Comparison

	Quantitative Methods	Qualitative Methods
Goal	Tests hypotheses Determines: whether (benefits, risks harm) & how much	Generates hypotheses Determines: what, how, why
Type of Reasoning	Deductive	Inductive
Study Designs	Randomized clinical trials Epidemiologic data Close-ended surveys	In-depth interview Focus groups Field observation Document content analysis
Outcomes	Frequency distributions P-values Effect sizes	Thick descriptions and/or theoretical structure Domains and attributes of phenomena
Question Structure	Close-ended: Continuous (e.g. age) Ordinal (pain scale 1- 10) Dichotomous (yes/no) Categorical (strongly agree, agree, etc.)	Open-ended Semi-structured interview Focus Groups with trigger questions In-depth interviews
Analyses	Univariate statistics Measures of association Multivariate statistical modeling (e.g. regression)	Grounded theory – open and axial coding Template coding Inter-rater reliability - kappa statistic
Sampling	Probability based -- designed to permit clinical and statistical significance based on frequency of observed outcomes	Theoretically based – designed to include observations from relevant and comprehensive pool of informants
Sources of Bias	Measurement selection Measurement error Reported via confidence intervals and statistical significance Most strongly associated with item <i>validity</i>	Investigator coding and interpretation Reported via kappa statistic of agreement beyond chance Checked by informant review of investigator interpretations Most strongly associated with item <i>reliability</i>

Qualitative Methods Appraisal Form

Citation:

Is the qualitative research relevant?	
Are the results credible?	
Is there a specific qualitative research method cited?	
Was the choice of participants or observations explicit and comprehensive?	
Were research ethics approval obtained?	
Was data collection sufficiently comprehensive and detailed?	
Were the data analyzed appropriately and the findings corroborated adequately?	
What are the results?	
What are the results?	
How can I apply the results to patient care?	
How does the study offer helpful theory?	
Does the study help me to understand the context of my practice?	
Does the study help me to understand social phenomena in my practice?	

Adapted from McMaster Evidence-based Clinical Practice Workshops and Users' Guide to the Medical Literature 3rd Ed.

Clinical Decision Analysis Appraisal Form

Citation:

Is this a newly derived instrument? (Level IV)	
Was validation restricted to the retrospective use of statistical techniques on the original database? (If so this is a Level IV rule). If so, consider the following standards for initial development of a decision rule.	
Were all important predictors included in the derivation process?	
Were all important predictors present in significant proportion of the study population?	
Does the rule make clinical sense?	
Has the instrument been validated? (Level II or III) If so, consider the following:	
Did validation include prospective studies on several different populations from that used to derive it (II), or was it restricted to a single population (III)?	
How well did the validation exercise meet the following criteria? Were the patients chosen in an unbiased fashion and do they represent a wide spectrum of severity of disease?	
Was there a blinded assessment of the criterion standard or outcome event (or was the outcome all-cause mortality) for all patients?	
Was there an explicit and accurate interpretation of the predictor variables and the actual rule without knowledge of the outcome?	
Was there 100% follow-up of those enrolled?	
How powerful is the rule (in terms of sensitivity and specificity; likelihood ratios; proportions with alternative outcomes; or relative risks or absolute outcome rates)?	
Has an impact analysis demonstrated change in clinical behavior or patient outcomes as a result of using the instrument? (Level I)	
How well did the study guard against bias in terms of differences at the start (concealed randomization, adjustment in analysis) or as the study proceeded (blinding, cointervention, loss to follow-up)	
What was the impact on clinician behaviour and patient-important outcomes?	

Adapted from McMaster Evidence-based Clinical Practice Workshops and Users' Guide to the Medical Literature 3rd Ed.

Clinical Practice Guidelines Appraisal Form

Citation:

Is the clinical question clear and comprehensive?	
Is the recommended intervention clear and actionable?	
Is the alternative clear?	
Were all of the relevant outcomes important to patients explicitly considered?	
Was the recommendation based on the best current evidence?	
Are values and preferences associated with the outcomes appropriately specified?	
Do the authors indicate the strength of their recommendations?	
Is the evidence supporting the recommendation easily understood?	
For strong recommendations, is the strength appropriate?	
For weak recommendations, does the information provided facilitate shared decision making?	
Was the influence of the conflict of interests minimized?	

Adapted from McMaster Evidence-based Clinical Practice Workshops and Users' Guide to the Medical Literature 3rd Ed.

Economic Analysis Appraisal Form

Citation:

Are the results valid?	
Did the recommendations consider all relevant patient groups, management options, and possible outcomes?	
Did the investigators adopt a sufficiently broad viewpoint?	
Are results reported separately for relevant patient subgroups?	
Is there a systematic review and summary of evidence linking options to outcomes for each relevant question?	
Were costs measured accurately?	
Did investigators consider the timing of costs and outcomes?	
What are the results?	
What were the incremental costs and effects of each strategy?	
Do incremental costs and effects differ between subgroups?	
How much does allowance for uncertainty change the results?	
How can I apply the results to patient care?	
Are the treatment benefits worth the harms and the costs?	
Can I expect similar costs in my setting?	

Adapted from McMaster Evidence-based Clinical Practice Workshops and the Users' Guide to the Medical Literature 3rd Ed.

Qualitative Example Teaching Package

Qua-Wars:

Qualitative Methods Unit – a comparison of quality and quantity

Learning Objectives:

1. To gain an appreciation for the role of qualitative research in health care.
2. To learn to distinguish qualitative from quantitative methods
3. To learn to appreciate and apply their results of qualitative papers including:
 - Determining the validity of the results
 - Interpreting the results
 - Applying the results to your patients

Problem-Based Educational Package:

1. Read the attached clinical scenario.
2. Define the clinical question and outline a strategy to find the evidence.
3. Read the attached JAMA articles about qualitative research in health care.
4. Review different teaching options based on the needs of your learners and the time available.
5. Critically appraise the attached paper
6. Decide whether the diagnostic test will help you in your decision making for your patient.

Enclosed Materials:

1. Steinhauser, KE et al. In search of a good death: observations of patients, families and providers. [Annals of Internal Medicine 2000; 132: 823-832.](#)
2. Giacomini, M et al. Users' Guides to the Medical Literature. XXIII. Qualitative research in health care, A. Are the results of the study valid? [JAMA 2000; 284 \(3\): 357-62.](#)
3. Giacomini, M et al. Users' Guides to the Medical Literature. XXIII. Qualitative research in health care, B. What are the results and how do they help me care for my patients? [JAMA 2000; 284 \(4\): 478-82.](#)
4. Strategies for Teaching Qualitative Methods
5. Critical Appraisal Worksheet
6. Steinhauser, KE et al. Factors considered important at the end of life by patients, family, physicians and other care providers. [JAMA 2000;284 \(19\):2476-2482.](#)

Clinical Scenario:

Mr. W is a 67-year-old man with a history of colon cancer 20 years ago that was treated at that time with surgery and chemotherapy. He presents to the ER with a 6-week history of difficulty swallowing and post-prandial emesis. He reports vomiting of undigested food and has lost 40 lbs in the past few months. He denies odynophagia. Mr. W also has a 40 pack-a-year smoking history and is a former heavy alcohol user, having quit several years ago. Physical Exam: Temp 99.3 BP lying down 138/70 P lying 80 with sitting up BP drops to 120/60 and pulse increases to 102. Mr. W is cachexic and has bitemporal wasting. His mucous membranes are dry. He has no LAN. He has decreased breath sounds at the right base. Heart exam is without murmurs or rubs. Abdomen, extremities, and neurologic examinations are normal.

This man is admitted to your service and over the next two days, Chest CT reveals marked thickening of the esophageal wall with near obliteration of the esophagus. EGD revealed an obstructive lesion with occlusion that was too tight to allow the endoscope to pass, however, a biopsy is done.

The team discusses the overwhelming probability of esophageal cancer with Mr. W while awaiting the results. Thus far, he has refused to discuss his hospital course with his family because he does not wish to worry them until he has an answer for sure. He was separated from his wife and does not maintain contact with her. He was in touch with his children, but they were all grown and he did not live near them. He openly admits, even without a clear diagnosis that if he has cancer, he would like to avoid toxic therapies if they will not be able to cure him. He reports that his inability to swallow is the single most difficult thing for him and that the quality of his life might be improved if he could swallow.

The student on the team voices helplessness at the situation. "It is clear that he is going to die and there is not much we can do for him." You suggest that, if he is going to die, we could try to offer him a "good" death. He responds, "a good death- what can be good about death?" You recall that you recently read a qualitative study that might help them to try to gain insight into the values and attitudes that surround death. They ask whether a qualitative study is one that is about quality. You recognize a teachable moment and offer to bring the paper the next day for the team to review. The paper you have in mind is:

In search of a good death: Observations of Patients, Families, and Providers. [Ann Intern Med. 2000;132:825-832](#)

Strategies for teaching About Qualitative Research

The Teaching Opportunity:

Background:

You have now identified a teachable topic for your team- you need to assess how much time you would like to spend on qualitative research methods and on the review of this paper. Each of the following 4 tips would lead up to a critical analysis of the qualitative study you have selected.

1. 15-minute tip:

You might begin by asking your group whether they can define and compare the characteristics of Quantitative vs. Qualitative Research with respect to the following categories. You can use the information in the User's Guides to help clarify the issues.

Table 1: Compare and Contrast Quantitative and Qualitative Research

What kind of information is being gained?	Quantitative Research	Qualitative Research
Type of Reasoning?		
Methods (Types of study designs used)		
Product of the work (i.e. what will be reported in your results section?)		
Measurements and Questions (Open vs Close ended)		
Statistical Considerations		
Sample Size Issues		
Sources of Bias		

2. 30-minute tip:

You might follow the exercise using table I (above) with a discussion surrounding Mr. W. Specifically ask, what are the things that we can do to improve the quality of Mr. W's time prior to his death? Have half of the group come up with Quantitative Outcomes and the other half come up with Qualitative Outcomes.

Quantitative Outcomes	How you would measure them	Qualitative Outcomes	How you would measure them

3. 60-90 minute workshop: Experiential learning about Qualitative Research

If you want to try something a little bit different, you could try an exercise that would really drive home the differences between the two research methodologies. Have your groups actually design and perform an experiment trying to clarify the following research question.

Research Objective: In search of a Good Breakfast: Observations of hungry medical residents—the Donut Trial.

Directions: Groups separated into qualitative and quantitative teams. You can use a coin toss to randomly assign people to their groups.

Materials: Two boxes of donuts, identical in content- one box per each team. (cookies or other food items would work as well as long as you have two identical sets)

Goal: Each team has 20 minutes to use their methodology to describe the donut breakfast in their group. The winning team will be that one which comes closest to “truth” about the breakfast provided as determined by an unbiased judge (to be picked by you!). Validity will be judged as the *truthful correspondence of results with an objective reality.*

The qualitative team must use open-ended evaluations, inductive reasoning and describe the important characteristics of the breakfast. They should use their data to generate hypotheses and may be experiential.

The quantitative team must use close-ended evaluations, deductive reasoning with hypothesis testing. Their hypotheses should be based on their prior experiences concerning what characteristics would be important. Attempts must be made to avoid bias. They must be able to apply quantitative statistics to their measurements (not actually do the statistics, just be able to!)

4. 60-minute tip:

A different approach would be to look at two papers that address similar questions with very different methods. Along with the qualitative paper from Annals of Internal Medicine, give out the paper from JAMA by the same authors (included in your packet). The JAMA paper reports on the quantitative work that followed the qualitative work. Ask the learners to read both papers and discuss their different strengths. Discuss why the second paper in JAMA could not have been done without the groundwork provided by the qualitative piece.

Assessing Validity: Is there a truthful correspondence of results to a presumed “objective reality.”

Methodologic rigor:

1. Is the study designed to address its research question and objectives appropriately?
2. Methods section: should include, participant selection, methods of data collection, comprehensiveness of data collection, procedures for analyzing data and corroborating findings.

User’s Guide to the Medical Literature: Critical Appraisal Worksheet for Qualitative Research:

Citation: Steinhauser, KE. et. al. In search of a good death: observations of patients, families, and providers. [Ann Int Med. 2000;132:825-832](#)

Validity Criteria	
What is the Research Question?	
Were participants relevant to the research question? Was participant selection well-reasoned?	
Were the data collection methods appropriate for the research objectives and setting? (Field observation, interviews, document analysis)	
Was the data collection comprehensive enough to support rich and robust descriptions of the observed events?	
Were the data appropriately analyzed and the findings adequately corroborated?	
Study Results	
What are the results of the study? How evocative and thorough is the description?	
Applicability to patient care	
How do the results of this study help me care for the patients? Does this study help me understand the context of my practice?	
Does this study help me understand my relationships with my patients and their families?	

Adapted from McMaster Evidence-based Clinical Practice Workshops

Qualitative Research Example ANSWERS

The clinical Question: In patients with **terminal** diagnoses, are there ways in which health care providers can help them prepare for **death**?

The search strategy: Using the key terms and concepts from the clinical question, you first type in “terminal” and your search engine will map you to the MESH heading *Terminal Care*. You will want to “focus” your search because the article is MAINLY about terminal care. Similarly, you will enter “death” and your search engine will map you to several choices, the best one being the MESH heading *Attitude to Death*. You will want to “focus” that one as well. Combining these two terms gives you 661 articles, so next you will want to use a methodologic filter (a set of search terms meant to select papers that use a particular methodology.) For qualitative studies, you want to find articles that are either interview studies or focus groups. Because the MESH heading for *focus groups* is underneath Interviews in the subject heading tree, if you “explode” interviews you will get all the articles indexed to interviews as well as focus groups.

The search:

1. *Terminal care/ 6124
2. *Attitude to Death/ 3352
3. 1 and 2 661
4. EXP Interviews/ 7400
5. 3 and 4 9

The paper: In search of a good death: Observations of Patients, Families, and Providers. *Ann Intern Med.* 2000;132:825-832

Notes:

* denotes ‘focus’; EXP denotes ‘explode’

Comments on the teaching tips:

General Comments: Most of us are much more attuned to quantitative methods and have little intuition for qualitative ones. Thus, each of the tips makes use of comparison to help clarify the unique properties of qualitative research. In addition, each tip is structured to drive home the point that qualitative research is **defined by the methodology, NOT by the research question or topic of study**. Thus, one can use qualitative or quantitative methods to answer any question whether it be about the quality of dying or breakfast foods!

Tip 1: A completed version of table I is included on the next page.

Tip 2: As noted in the general comments, you can address many of the same questions using each methodology. For example, spiritual awareness can be examined using in-depth interviews with patients and their families and also be assessed through a closed-ended survey tool.

Tip 3: If you set up each group with a clear view of what the methodologic “rules” are, you will find this an incredibly enlightening exercise. It is likely that the group will agree that the qualitative methodology produces more useful information than the quantitative methodology- in addition, the quantitative researchers each get to eat the donuts whereas, depending on the experimental design, the quantitative group may or may not eat! You must be sure that the quantitative group tests a particular hypothesis.

Tip 4: The nice thing about the comparison of these two articles is that they relate to the very same topic, by the same authors but use different methods. Each paper is methodologically strong but contributed a different

Table 1: Compare and Contrast Quantitative and Qualitative Research

	Quantitative Research	Qualitative Research
What kind of information is being gained?	Tests well-defined hypotheses. Determines, whether benefits, risks, harm) or how much (prognostic, diagnostic value)	Insight into emotional and experiential phenomena in health care; interpretive research. (Generates hypotheses) Determines what, how & why
Type of Reasoning?	Deductive	Inductive
Methods (Types of study designs used)	<ul style="list-style-type: none"> - Clinical Trials (RCT for therapy / harm) - Epidemiologic Data (Cohort for prognosis) - Diagnostic Test Study Close-ended survey 	<p>Interviews (semi-structured, in-depth, individual and focus groups)</p> <p>Field Observation (direct vs indirect)</p> <p>Document Analysis (journals, charts, correspondence)</p>
Product of the work (i.e. what will be reported in your results section)	Explicit measured outcomes	Narrative, story that describes and explains social phenomena
Measurements and Questions (Open vs Closed-ended)	Closed	Open
Statistical Considerations	Can apply quantitative statistics	Cannot apply quantitative statistics, results are open to interpretation
Sample Size Issues	Explicit numbers of patients are targeted to reach clinical / statistical significance based on the frequency of the observed outcome	Adequately, in-depth observations; Theoretical saturation of themes (informational redundancy) breadth of observations and depth of each observation made; data collection until you produce data in enough detail to represent the experience
Sources of Bias	<p>Measurement selection</p> <p>Measurement error</p> <p>Reported via confidence intervals and statistical significance</p> <p>Most strongly associated with item validity</p>	<p>Investigator coding and interpretation</p> <p>Reported via kappa statistic of agreement beyond chance</p> <p>Checked by informant review of investigator interpretations</p> <p>Most strongly associated with item reliability</p>

Assessing Validity: Is there a truthful correspondence of results to a presumed “objective reality.”

Methodologic rigor:

1. Is the study designed to address its research question and objectives appropriately?
2. Methods section: should include, participant selection, methods of data collection, comprehensiveness of data collection, procedures for analyzing data and corroborating findings.

User’s Guide to the Medical Literature: Critical Appraisal Worksheet for Qualitative

Research:

Citation: Steinhauser, KE. et. al. In search of a good death: observations of patients, families, and providers. [Ann Int Med. 2000;132:825-832](#)

Validity Criteria	
What is the Research Question?	What are the attributes of a good death, as understood by various participants in end-of-life care?
1. Were participants relevant to the research question?	Yes. Physicians, nurses, social workers, chaplains, hospice volunteers, patients, and recently bereaved family members were all involved.
2. Was participant selection well-reasoned?	<p>Yes. The participants involved the full spectrum of persons involved with end-of-life care. In addition, participants were recruited from a variety of settings including a private, tertiary care hospital, a VA hospital, a local community hospice in Durham, NC</p> <p>Non-physician providers: recruited from convenience samples generated by e-mail and departmental advertising.</p> <p>Physician recruitment: stratified by level of appointment, then randomized the lists and recruited in order ensuring that the overall group represented each career level.</p> <p>Patient recruitment: HIV and oncology patients recruited by telephone. Stratification for ethnicity</p> <p>Family members: stratified random sample of recently bereaved relatives of veteran patients who had died 6 months to 1-year prior.</p>
3. Were the data collection methods appropriate for the research objectives and setting? (Field observation, interviews, document analysis)	The methodology included focus groups and in-depth interviews. Focus groups included 6 to 8 participants and were stratified by ethnicity with trained facilitators of the same race as the participants in each group.
4. Was the data collection comprehensive enough to support rich and robust descriptions of the observed events?	<p>Appendix describes protocol for focus group discussion.</p> <p>Particular care was taken to ensure reliability and exhaustiveness of the data collection and analysis.</p> <ol style="list-style-type: none"> 1. Exhaustiveness: focus groups were conducted until same themes were repeated and no new themes emerged. 2. After repeatedly analyzing the focus group transcripts, in-depth interviews were conducted with the most talkative and the quietest member of each group. This was done to ensure that no new themes arose and also to give the more quiet participants the chance to voice possible silent, but dissenting viewpoints.

3. Were the data appropriately analyzed and the findings adequately corroborated?	<ol style="list-style-type: none"> 1. Focus groups and interviews were transcribed from audiotapes. 2. Constant Comparisons Methodology was used for coding. Using this method, 4 investigators independently read examples of the transcripts and analyzed them for recurrent themes. Summaries were then compared for agreement. 3. 70 attributes of a 'good death' were identified and collapsed into 6 broad domains. 4. 4. Illustrative quotes were included
Results	
What are the results of the study? How evocative and thorough is the description?	<p>The study describes the 6 domains are important components of a good death:</p> <ol style="list-style-type: none"> 1. Pain and symptom management 2. Clear decision-making 3. Preparation for death 4. Completion 5. Contributing to others 5. 6. Affirmation of the whole person <p>Each domain is described and illustrated by a quote from the focus groups in interviews that captures the spirit of the domain.</p> <p>Of note, the 6 domains contain four that have been previously incorporated into palliative care paradigms, however, the importance of contributing to others and affirmation of the whole person were previously unrecognized in their role at the end of life. This is a perfect example of how a qualitative, open-ended methodology can expand our paradigms of care.</p>
Applicability to patient care	
How do the results of this study help me care for the patients? Does this study help me understand the context of my practice?	<p>These data enhance awareness to those domains that are important in end-of-life care. Attention to these aspects of the dying patient's experience will enhance the quality of their death.</p> <p>Yes.</p>
Does this study help me understand my relationships with my patients and their families?	<p>Yes. These data provide a road map to assist in the navigation of end-of-life issues.</p>

Adapted from McMaster Evidence-based Clinical Practice Workshops

Resolution of the clinical scenario:

Mr. W's diagnosis turned out to be stage IV squamous cell esophageal cancer. The team had discussed his end of life wishes with him and he did not wish to be resuscitated. Because he wanted to be able to swallow, he did elect palliative XRT and chemotherapy with 5-FU/Cisplatin. He had a PEG placed for nutrition. The patient died 6 weeks after he presented to the ER. His death was in the hospital when he was receiving chemotherapy. The same team of residents who originally admitted him was caring for him. He began to experience delirium, then rather quickly and unexpectedly died. He was on morphine for pain control, he was not resuscitated and his family did not wish to have an autopsy. Following this patient's death, the team expressed regrets that they had not discussed his spiritual needs and that he died without a clear mental status.

Teaching Strategies

Teaching Strategies Introduction: Teaching Tips and Materials

Notes about learning and teaching this section's **Teaching Tips and Materials**:

1. **Filling the Tool Bag:**

This is a series of teaching tips and strategies for teaching EBM that have been compiled from our collective teaching experiences, i.e., from the successes and failures of experienced EBM educators in Canada and the U.S. In addition to the actual tool-bag table is an article entitled "EBM Package Writer Suggestions" which outlines helpful hints on how to select articles and craft EBM sessions aimed at educators who wish to write and teach EBM sessions or workshops.

2. **Ridicularium Exemplario:**

This is a teaching strategy that can help learners when you or they get 'stuck' in the details.

3. **Critical Appraisal Sheets:**

This "teaching tips" section also includes an extra, blank set of critical appraisal sheets for all of the topics covered (for use in the future and for copying, etc.). In addition, you will also find critical appraisal sheets for future use with topics not covered at this year's conference.

4. **Curriculum Planner Workbook for Residency Training:**

This workbook addresses the particular needs and objectives of residency education although most of the principles, guidelines, and strategies presented here are applicable to almost any teaching situation.

Tips for Creating Examples: Writer Suggestions to Consider

Background:

Writing effective teaching packages for evidence-based medicine is challenging. Teaching settings, venues, audiences, and experience levels of teachers and students vary widely. Nonetheless, there may be elements of a teaching package and approaches that increase the likelihood of a successful package. In order to explore methods for writing successful packages, we surveyed experienced teachers of evidence-based medicine who have had several years of experience with package writing. Specifically, we asked them to identify features of past packages that have produced success as well as features of past packages that have created stumbling blocks for learners. The following summary of suggestions may help guide you in writing your own teaching packages.

General Summary of findings:

Respondents agreed that the key to a successful package is a *methodologically strong paper as it is applied to a clinically interesting case*. In addition, it was felt that there should be adherence to certain 'rules' in terms of which methodology is highlighted in each package as well as the spectrum of material covered in each. Those package writers who responded felt overall that we need to *simplify* the packages and perhaps stick more closely to simply providing the clinical case, possibly teaching settings and the critical appraisal materials (including, of course, application).

Summary of particular points:

1. Format and material covered should be consistent. Suggestion: Package writers might consider the following *outline* for the flow of each teaching package.
 - Clinical Case / Teaching setting(s)
 - Clinical Question Formation
 - Brief comment on acquiring the evidence
 - Summary of material in package, section(s) of Users' Guides book where relevant methodological discussion is to be found
 - Critical Appraisal sheet filled in with application addressed in this context
2. Certain teaching packages should be consistent in study design used. Suggestions:
 - Therapy: RCT
 - Harm: Cohort or Case Control Study to allow the participants the opportunity to learn about and practice these study types
 - Meta-analysis: summary of therapy trials, as opposed to other types of questions
 - Prognosis: Case Control or (much more frequently) Cohort methodology (can be in the context of an RCT)
 - Diagnosis: Prospective cohort with comparison to a reference standard
3. General Strategies that have produced *successful* teaching packages in the past:
 - Papers with clear, transparent, excellent methods sections
 - Clinical cases that are engaging, that may provide a new perspective that clinicians were not aware of or that provide points for interesting consideration regarding application of evidence

4. General Strategies that have produced *difficult* teaching packages in the past:
 - Poor methodology of the paper
 - Unclear or incomplete methods sections
 - Beware of papers that have their methods described in another paper
 - Cases or papers that are too complex
 - Uninteresting or irrelevant clinical problems
5. *Fun* suggestions that might be tried in the future:
 - Inclusion of expected stumbling blocks and troubleshooting strategies particular to the specific package
 - Cases that use multiple versions of the same evidence (e.g. ACP journal club summaries as well as the entire article)
6. Feedback on *Diagnosis* Teaching Packages:
 - The paper should provide enough data to calculate or extract multiple levels of Likelihood Ratios (LRs).
 - Papers with dichotomous outcomes may not illustrate the power of LRs.
 - It is important to highlight the great impact of patient values on the application of test results.
 - It is valuable to include discussion of test threshold and action threshold.
7. Feedback on *Systematic Review / Meta-analysis* Teaching Packages:
 - The paper should focus on therapy and summaries of RCTs.
 - Forest plots are very useful for teaching concepts including heterogeneity.
8. Feedback on *Therapy* Teaching Packages:
 - It is generally necessary to choose a positive trial with at least one dichotomous variable. Otherwise, there isn't an opportunity to practice RRR, RD, NNT.
 - As one of the fundamental packages and as therapy is the most prevalent kind of paper in the literature, it makes sense to keep this one 'timely' and on the forefront of emerging therapies.
 - Applicability and generalisability should always be addressed.
 - Ideally, it will be easy to identify sub-groups at different baseline risk to get an accurate notion of the baseline risk and to facilitate using baseline risk and RRR to calculate NNT.
 - It might be fun to have a low-risk group that would lead one to question the treatment (moving the threshold NNT) in the setting of an appreciable harm/cost to balance the benefit.
9. Feedback on *Guideline* Teaching Packages:
 - Repeated difficulty has been linked to the very lengthy nature of many good guidelines (most are 20-50 pages or more!).
 - To get around lots of reading one might:
 - a) Direct learners to key parts of the methods and results instead of the entire guideline.
 - b) Select one recommendation in the guideline and focus on that one.
 - c) Use resources and summaries available on the web (e.g. www.guidelines.gov).

Ridicularum Exemplario

The Challenge: To keep learners focused on the knowledge or skill set that you want to address and to avoid digressions over clinical passion...

Example: *Has this ever happened to you?*

Scenario: You were asked to come to a meeting of nurses at your hospital to help them generate some excitement for EBM. Specifically, you want to help them get excited about clinical question formation. They have little to no background in EBM. You present a case of a hospitalized patient with delirium because most of them are inpatient nurses and you wanted the case to 'hit home' and be relevant to them. As you begin trying to draw clinical questions out of them, they begin arguing about the clinical scenario. They ask you endless questions about the clinical case, the providers involved, the color of the room that the patient was in, the size of the hospital gown...When it is clear that they cannot move past the 'facts' of the case, you sigh and try a different approach.

Teaching goal: to practice skills in clinical question formation

Stumbling block: the details of the case generated excitement but got in the way of your message

Possible alternative strategy: Ridicularum Exemplario

Core components of Ridicularum

- Derive an example that is completely ridiculous, but sets the stage for a discussion of the curricular points you want to make.
- The example can be non-medical or medical but it must avoid any link to reality to be effective.
- Can be used for any teaching quest
- Can be incredibly engaging and lots of fun!!

Ridicularia from Durham:

- The DONuT: (The **D**uke **O**bservational **N**utrition **T**rial)-
Goal: to teach about qualitative methods. Residents are randomized to a qualitative methods arm or a quantitative methods arm. Once in randomly assigned groups, they are given a 'grant' from Dunkin' Donuts to design a trial to identify the qualities of a good donut breakfast. They **MUST** design the trial using the methodologies associated with their randomized groups.
- The Parking Ticket:
Goal: to teach the principles of decision analysis. A scenario about a parking decision regarding parking in a nearby illegal spot vs. the farther away pay parking lot.
- The Dancing Ballerinas:
Goal: to teach principles of risk. A scenario about risks involved with wearing different colors of ballet shoes and an intervention that can change the color of the shoe in a proportion of ballerinas.
- The Drive Home:
Goal: to teach principles regarding prognosis. A scenario regarding a 'spirited discussion' between a husband and wife on the way home from a dinner party.
Discussion surrounds risks involved with the driving behaviors of one of the spouses. (This scenario is based on a real interaction but the names have been changed to protect the innocent.)
- Messages from Mom:
A series of trigger audio clips of a rather overprotective mom regarding the health and safety of her daughter. The goal is to practice clinical question formation in the context of these examples.

Filling the Tool Bag Strategies for making it fun and effective

Teaching Objectives	Setting	Strategy
Overall Objectives		<ul style="list-style-type: none"> Always define your goals in advance and discuss them at the beginning of the session. Limit yourself to three major goals per session (more will be lost, and you may risk losing all of them).
Clinical Relevance		<ul style="list-style-type: none"> Begin and end with a patient case / clinical question.
	Keep the learners involved.	<ul style="list-style-type: none"> Ask open-ended questions. When someone asks a question, turn it back to the group, i.e., “what does the group think?” or “can anyone help out here?” (This also buys the tutor some time, in case the answer isn’t immediately apparent to the tutor!)
Incorporation of Values into Decision-Making		<ul style="list-style-type: none"> Openly discuss the portion of decision-making that remains intuitive, emotional, sensitive to the needs of the patient and the community in which they live.
Physical Needs	Recognize the limits of your learners’ tolerance.	<ul style="list-style-type: none"> Take time for a stretch. Attend to food needs at all times. Cookies are therapeutic.
Emotional Needs	Effective learning requires an emotionally-safe environment.	<ul style="list-style-type: none"> Make sure everyone knows it is okay not to know! Make sure everyone knows it is okay to disagree (agreeably)! Be open about your own limitations! Look for opportunities to compliment and praise. Call “time-outs” when the group dynamic becomes tense. Ask the group what is happening with the process, and then try to return the focus to the problem/case.
	Capitalize on disagreement.	<ul style="list-style-type: none"> Try to incorporate the rest of the group into the discussion. Seize the right opportunity for wrap-up or closure.
Using Examples / Case Scenarios	Pre-test probability	<ul style="list-style-type: none"> Use cases in order to capture very low-risk patient, very high-risk patient and very ‘toss of a coin’ risk patient.
	Diagnostic Tests	<ul style="list-style-type: none"> Use extreme examples of cases to make people commit to a pre-test probability. Use the examples to define cases that are extremely low pre-test probability, extremely high pre-test probability, and the middle cases.
Language Pitfalls		<ul style="list-style-type: none"> Don’t use jargon / use simple plain language. Ask those learners who use jargon to explain the term(s) to the rest of the group.
	General Strategies	<ul style="list-style-type: none"> Vary the pace of your session by taking time out to give specific tasks or skills-practice to the group members. Break your group into smaller working groups of 2-5 people.

		<ul style="list-style-type: none"> Be very clear about what you want them to do (e.g., “assess the therapy validity criteria for this paper” or “take 5 minutes to review the methods and describe the patient population”).
User-Friendly Statistics		<ul style="list-style-type: none"> Emphasize the difference between “statistical significance” and “clinical importance.”
User-Friendly Statistics	Calculations: NNT, LR	<ul style="list-style-type: none"> Set the exercise up properly: <ol style="list-style-type: none"> Set up the importance in the big picture. Model the calculation (show them one). Ask them to do another permutation. Return to the original importance of the calculation. Give defined small tasks and break into groups of 2 or 3 to do specific calculations.
Teaching Definitions	Confidence intervals, likelihood ratios, sensitivity, specificity etc.	<ul style="list-style-type: none"> Try several different ways of defining the same thing - coming from different viewpoints Relate it to a scenario or example so we can put the definition into a framework.
User-Friendly Statistics	Learning to Love a Likelihood Ratio	<ul style="list-style-type: none"> Take home points: <ol style="list-style-type: none"> You can utilize likelihood ratios for a range of values for a given diagnostic test – i.e. it is not a (+ or -) dichotomous measure. Sensitivity / Specificity are properties of the test, Positive, and Negative. Predictive value is properties of a test in a population, LR allows you to apply the test directly to an individual patient. The likelihood moves you from a pretest to a post-test probability. You must estimate pre-test probabilities first and acknowledge the uncertainty that goes along with that. You don’t have to draw the 2x2 table if you don’t want to do calculations and just want to talk about LR. If you want to calculate, it is hard to do without the 2x2.
Directed Engagement of Learners		<ul style="list-style-type: none"> Assign your learners to a point of view, a role or a specific task. Clinical Practice Guideline – Randomize one-half of the room to ‘love them’ and one-half of the room to ‘hate them.’
Silence	Groups or individuals who will not participate	<ul style="list-style-type: none"> 16-second rule: Refrain from jumping in to fill the silence yourself! (May require longer for cultures in which participation is less accepted; may require shorter for people from New York!)
Discussion dominators		<ul style="list-style-type: none"> Use “time-outs” when someone is dominating the discussion or ‘knows it all.’ Ask the group members to talk about individual responsibilities (for loud ones to lighten up and quiet ones to contribute more).

Using the Blackboard		<ul style="list-style-type: none"> • Plan in advance what you will do. • Put up one thing at a time and orient the group to what you are writing up there. • Have someone else write on the board so that you can focus on teaching and to optimize engagement. • Have the other learners direct their peer at the board in what to do.
Using Handouts	Using Tables and Figures	<ul style="list-style-type: none"> • Hand out only what you need. • Give brief orientation to the table. • Be specific in your direction of what you want people to see from the table.
	Reinforcing and providing resources for home	<ul style="list-style-type: none"> • If you hand something out, people will read it instead of listening to you; hand out take home papers at the end. • Do write down formulas and calculations if you believe in their importance. • Tell learners at the beginning that you will provide a handout so that they can focus on participating rather than taking notes.
Issues of Time Management	How to deal with questions that come up that you don't have time to answer?	<ul style="list-style-type: none"> • Answer quickly. • Canvas the Group, diagnose your learners. • Return to the "Parking Lot."
	Be Realistic	<ul style="list-style-type: none"> • You always have less time than you think you do. • Juicy issues are fun, but also juicy—they take time! Budget for it. • Stop from time to time to synthesize/summarize – for emphasis and to check in with learners.
	Trim the Fat	<ul style="list-style-type: none"> • Clearly define your teaching goals so that you can differentiate what you must have from what you may have from what should be cut.
	Save time for closure.	<ul style="list-style-type: none"> • Come to closure about the article and the clinical scenario. • Closure does not mean "unanimous agreement."
	Clinical Practice Guidelines, Decision and Economic Analysis	<ul style="list-style-type: none"> • Perspective is a key teaching point for each of these methodologies. • Divide into groups and assign perspectives (the managed care plan, the patient/family, the doctor's office, the hospital, society).
		<ul style="list-style-type: none"> • HAVE FUN! If you enjoy what you do, your learners will too.

Feedback: 6T's Teaching Tips

(Figurski, Patel, Keitz, Cook, EBM Workshop 2005)

Objective: To provide a touchstone to plan and evaluate each teaching session

Teaching Utility:

- 1) Provides tips to help plan teaching sessions (but is not exhaustive)
- 2) Provides framework for session evaluation (if they get too detailed)
- 3) Symbolizes group culture (can add logos for fun)
- 4) Can be modified (this used to be 4Ts in June 2005)

The 6 T's	Observations and Suggestions
<u>T</u> ime management (before & during)	
<u>T</u> eamwork (ensure engagement)	
<u>T</u> ools (use them)	
<u>T</u> riage (decide what you can't cover)	
<u>T</u> one (respectful, safe)	
<u>T</u> ake home message(s) (obligatory!)	
Other <u>T</u> hings...	

Curriculum Planner Workbook

Building a curriculum in evidence-based clinical practice

Background: What is a curriculum?

The **definition of a curriculum** is simply a **planned educational experience**. It is a systematic planning approach that some educators use to help them articulate, achieve and evaluate educational goals.

Most medical education curriculums are based on a behaviorist model of education. In this model, there is the identification of specific learning objectives and an emphasis on the acquisition of competency of various knowledge and performance tasks. Increasingly EBM enthusiasts are aware that this approach is necessary but not sufficient to help our learners emerge as Evidence- Based practitioners. There are lessons from other educational models that educators interested in the area of EBM should explore, such as the social learning theory or cognitive theory. For example, social learning theory emphasizes the importance of a collaborative learning environment in which new learners are mentored through interactions with role models.

A successful curriculum is one that addresses the unique needs and goals of your learners and educators in your own setting. The information that follows is meant as a resource to help stimulate thought and planning in your home environment.

Steps to building your own Curriculum

There are several models for the creation of an academic curriculum. The following is an example of some of the steps to take to start you on your way.

1. Problem ID and General Needs Assessment e.g. “The world needs Evidence-Based Medicine”
2. Needs Assessment of Targeted Learners
 - Identify your targeted learners e.g. the interns, the residents, the urology dept. etc.
 - Determine specific needs of your targeted learners e.g. to question build, to search, to critically appraise a therapy article, to incorporate patient values into decision-making
3. Formulation of Curricular Goals and Specific Measurable Objectives
 - Who will do how much of what by when?
 1. Who? e.g. my interns
 2. Will do? e.g. will search the medical literature
 3. How much? e.g. for seven clinical questions
 4. Of what? e.g. pertaining to therapy questions
 5. By when? e.g. by next Wednesday
4. Educational Strategies Lecture, workshop, skills practice, self-directed learning, role modeling
5. Implementation of the Curriculum which room, what time, which kind of cookies will you serve
6. Evaluation and Feedback If you evaluate, it implies importance and allows improvement

Putting together a curriculum specific for teaching evidence-based clinical practice at home

The following pages go through some steps and guideposts that may help you to design a curriculum for evidence-based clinical practice at your home institutions. The information is divided into the following sections:

- I. REMEMBERING THE BASICS
- II. THE EVIDENCE BASED PRACTICE COMPETENCY GRID
- III. QUESTION BUILDING
- IV. SEARCHING THE MEDICAL LITERATURE
- V. CRITICAL APPRAISAL STEPPING STONES
- VI. APPLICATION TO PATIENT CARE
- VII. TARGETED NEEDS ASSESSMENT

I. Remembering the Basics

In setting up your curriculum it is essential that you remember one of the golden rules of evidence-based clinical practice: evidence-based practice begins and ends with the patient. No matter how you put things together, the inclusion of clinical scenarios that are meaningful to the teachers and learners is essential to a successful curriculum. In planning your strategies, you should keep in mind the five steps that are linked together to promote the incorporation of best evidence into clinical practice. These steps are summarized as follows:

1. Question formulation derived from patient care
2. The selection of appropriate information resources and the identification of evidence from the medical literature
3. Critical appraisal (determining validity, evaluating the magnitude of results and determining applicability)
4. Returning to the clinical situation at hand to decide how to implement the evidence

II. The Evidence-Based Practice Competency Grid

The attached grid describes many of the specific knowledge, attitudes and skills that are necessary for the various steps in the evidence-based exercise. The grid can be used to identify jumping points for your curricular planning.

The grid is part of a work in progress by Scott Richardson, Mark Wilson, and Sheri Keitz. We encourage your feedback as you reflect on your learners' needs. Please pass along thoughts and comments on how we can improve the grid to Sheri Keitz (sheri.a.keitz@lahey.org)

III. Question Building

How to ask questions

Asking a clinical question that can be answered is one of the most important skills that you will teach your learners. A well-built clinical question derived from patient care is necessary to drive the subsequent steps in the process. The nature of the clinical question will drive the choice of information resources and focus an effective MEDLINE/PubMed search strategy. The kind of question that you ask (e.g. therapy vs prognosis) will determine the kind of research study that you will want to find (e.g. randomized controlled trial vs cohort study) as well as the critical appraisal skills that you will need to determine validity. Finally, the clinical questions will ultimately guide your use of the information retrieved when you decide whether that information is applicable to your individual case at hand. Given the central and essential role of the clinical question, the skills, attitudes and knowledge that relate to question building must be central and essential to any curriculum planning.

As with any skill, it needs to be modeled, taught, reinforced, and practiced. Learners must have feedback to help them understand why certain questions lead to fruitful searches and why some questions don't. Of equal importance, they must learn to prioritize which questions they should pursue with rigor and which questions should occupy less of their time. Without practice, feedback and prioritization skill our learners may become frustrated and disillusioned.

You may wish to dedicate some formal, academic time to a workshop dedicated to question building. In addition, many of us use question cards, such as the one enclosed in this Workbook to facilitate ongoing questioning. No matter how you choose to reinforce the importance of the clinical question, you will likely need to combine several teaching strategies to get your learners tuned into this critical skill.

IV. Searching the Medical Literature

MEDLINE (via PubMed or Ovid) Strategy Assessment Tool

The purpose of this instrument is to break down the searching process into the fundamental parts and concepts. A teacher can use it to define elements of the curriculum that are necessary for teaching effective searching skills. A learner can use it to evaluate their own searches to determine whether or not they have used the appropriate concepts to search the Medline database. This tool was created at Duke University by Connie Schardt, MLS, Chris Cabel, MD and Sheri Keitz, MD, Ph.D. We are testing the utility of this tools and are eager for feedback. Please send comments to Sheri Keitz (sheri.a.keitz@lahey.org) and Sarah Cantrell (sarah.cantrell@duke.edu).

I. Getting Started-- Fundamental Information to get you started	
What is the Clinical Question?	
What are the key elements of the question?	
II. Primary Guides to Effective Searching:	
Did the Search Strategy address all of the key elements of the Search Question?	
Was the question divided into concepts and each concept searched separately?	
Were sets combined correctly (i.e. appropriate use of Boolean logic: AND, OR)	
III. Secondary Guides to Effective Searching:	
Were MESH terms used whenever possible?	
Were text words used when MESH headings were not available or appropriate?	
IV. Additional Strategies for fine tuning a search:	
Was methodologic filtering used? (Including clinical query filters, publication types, or MeSH that address research methods)	
Were subheadings used?	
Was text word truncation and adjacency used?	
Were appropriate limits applied?	

V. Critical Appraisal Stepping Stones

All Critical Appraisal exercises are not creating equally: Part I

When selecting which type of articles to teach and discuss, learners' levels of sophistication, background and experience must be taken into account. For all learners, the questions most commonly asked related to therapy, followed by diagnostic testing, etiology/harm, and prognosis. Therefore, it will be our goal to sure that the learners are comfortable with the following topics. (Note: a more detailed listing of competencies is presented in the competency grid.)

Essential Tools	Critical points of knowledge
1. Question Building	4 parts of a clinical question and question 'map'
2. Acquiring the evidence	Accessing the literature, MEDLINE (via PubMed or Ovid) searching, electronic resources
3. Therapy	Number needed to treat; Trisk/ Benefit Ratio
4. Diagnostic Testing	Likelihood ratios
5. Etiology/ Harm	Number needed to harm; Case-Control/ Cohort study methods
6. Prognosis	Bias; inception cohort; cohort study methods
7. Overview articles	Focused question, assessment of comprehensive search and study rigor
8. Clinical Practice Guidelines	Comprehensive look at options/ outcomes/ literature

Once learners are up to speed on those skills that relate to these topics, consideration can be made to moving on to other topics such as economic analysis and clinical decision analysis.

All Critical Appraisal exercises are not creating equally: Part II

When selecting specific articles for teaching and discussion, once again consideration must be made to the sophistication of the learners. For all learners, it is of critical importance that the articles be of clinical interest to them. For early learners, a more directed approach may be taken in ensuring that they select methodologically strong articles to optimize success for the critical appraisal exercise. When early learners select their own articles, there is the risk that many of the articles selected may be 'fatally' flawed and learners will not feel successful. It is important to avoid nihilism ("all articles are flawed; therefore, why should we do this?") In this setting, appropriate feedback and redirection are often necessary for naïve learners.

VI. Application to Clinical Care

Evidence-based medicine begins and ends with the patient or clinical situation. It is essential that issues of applicability be addressed for each paper reviewed. This allows discussion of the psychosocial context for care, issues of financial and / or social constraints, patient and societal values. Key concepts include 'generalizability', comfort with value-laden decision-making and strength of inference.

VII. Targeted Needs Assessment

Identification of resources and barriers

- I. Who are my learners and when and where can I teach them? Identify each set of learners (e.g. interns, residents, fellows etc.) you are targeting and the various settings you want to impact in your curriculum (e.g. ambulatory block rotation, ward service, morning report, journal club etc.) Be sure to identify the settings that are appropriate for each set of learners.

Learners:	Setting:

- II. Who are my colleagues and what can I convince them to do?

Colleagues:	Tasks that they can help with:

- III. What materials and resources do I require? Of those, what do I have and what do I need to get? Specifically consider faculty time, computer resources, space

Resource	Have it (Name it)	Need it (How can I get it)

- IV. What are the barriers to your success and how are you going to solve them?

Barrier:	Solution:

Evidence Based Clinical Practice Competency Grid

	Knowledge	Attitudes	Skills
Assess: Assess the patient and the clinical scenario	- Basic clinical skills (H&P) and disease-specific knowledge	- Acceptance of knowledge deficits - Interest in self-improvement and in increasing fund of knowledge	- Formulation of clinical assessment of the individual patient - Assessment of learning needs as a health professional
Ask: Clinical Question Formation	- The anatomy of a question - The Map for Clinical Questions (e.g. therapy vs diagnostics vs prognosis)	- Curiosity - Comfort with Uncertainty - Value active learning - Learn to sort through which questions are of greatest importance to you or your patients.	- Formulate a question - Identify it's "location" on the Map for clinical questions - Identify the research method that will best answer the question (e.g. RCT vs cohort) - Create a hierarchy of importance for which questions you will invest time and energy in
Acquire: Selecting and getting the evidence			
A) Searching the Medical Literature	- MEDLINE as a database - MeSH vs. Textword / Keyword Searching - Methodologic filtering	- Fear of the volume of available medical literature - Deal with aversion to technologies	- Tie key elements of the question to specific search strategies - Timely and efficient searching - Boolean Logic: (And / Or / Not) - Explode, Focus, Truncation, Limits and Subheadings
B) EBM Resources	- Awareness of alternative resources Assessment of evidence-based nature of resources	- Willingness to explore numerous apps and resources - Value efficiency	- Computer literacy - Informatics
Appraise: Critical Appraisal	- Practical clinical epidemiology (User's Guide to the Medical Literature) - Primary Guides vs secondary guides for validity - Fatal Flaws - Survival Statistics - Creating a hierarchy of evidence	- Address innumeracy - Promote readiness to challenge authority (Challenge them to be critical, don't accept it as it must be so) - Promote enthusiasm and avoid Nihilism	- Identify which article will answer your question - Apply these skills real time settings
Apply: Application of Evidence to Clinical Care	- Getting the individual patient Number needed to treat (NNT) or Number needed to Harm (NNH) - Going from pre-test to post-test probabilities (likelihood ratios) - Strength of inference	- The recognition that value judgments are implicit in every clinical decision and are being made all the time by physicians based on the MDs and patient's value systems - Comfort with making value-based recommendations	- Solicit patient preferences - Assess co-morbidity - Consider social support of patient - Assess where the patient's value system lies on the paternalism to technical continuum
Evaluation of Performance	- Understanding the elements of quality measurement and self-assessment	- Addressing reluctance to assess one's own behavior to identify areas for improvement - Readiness and willingness to change one's own behavior	- Measure / Assess - Intervene - Re-measure /Reassess

Sample Evidence-Based Medicine Curriculum

Building a Curriculum in Evidence Based Medicine

Example of a Curriculum Document

Example:

A plan for an Evidence-Based Medicine Curriculum at Duke University /Durham VA Medical Center

Background: The following example is a curriculum-planning document from Duke University, NC, and the Durham VA PRIME Program. At Duke and the VA, we have been running EBM workshops for the past 6 years. In planning for the next academic year, we wanted to increase the use small groups of residents working closely to create mentorship opportunities between senior residents and interns.

Please direct questions, thoughts or comments about this curriculum to Sheri Keitz

(sheri.a.keitz@lahey.org)

1. General Needs Assessment:
 - General Goals of an EBM Curriculum: To implement a structured series of workshops to provide the fund of knowledge and skills necessary to incorporate the best evidence in the care of individual patients
 - Current Workshop series is well received but not uniformly delivered to all Medicine House officers
 - We have limited faculty to teach the house officers- if we want to do more teaching, we will need the house officers to do it
 - Task: To Disseminate the Curriculum to all house officers over the course of the three-year residency program using the residents to teach and mentor each other
2. Needs Assessment of Targeted Learners
 - INTERN Goals:
 - To Instruct them in Question Building and Basic Searching Skills
 - To provide role models and examples of EBM in practice
 - SECOND YEAR RESIDENT Goals:
 - To provide them with the teaching and leadership skills
 - To provide them with content and skills base to practice and teach EBM
 - To pair them with faculty mentors to co-teach sessions on EBM
 - THIRD YEAR RESIDENT Goals:
 - To allow them to teach and lead sessions on EBM independently
 - To give them the opportunity to be role models and mentors for interns
3. Formulation of Broad Curricular Goals
 - Interns will be exposed to question building and searching skills in two large group sessions (60 minutes each) lead by faculty
 - Interns will practice the EBM exercise in workshops lead by SECOND YEAR RESIDENTS and THIRD YEAR RESIDENTS (60 to 90-minute sessions)
 - SECOND YEAR RESIDENTS will be paired with faculty mentors on their Ambulatory Block Time and each will co-lead one EBM session and participate in the sessions lead by their peers.
 - THIRD YEAR RESIDENTS will participate in learning teams with interns and each will lead one EBM session and participate in the sessions lead by their peers.
4. Specific Measurable Objectives
 - All workshops will focus on the Competencies needed for the Practice of Evidence-Based Medicine (See Table)
 - Objectives of all Workshops:

After completing this workshop, you should be able to:

1. Create a pertinent answerable question from a clinical case scenario
 2. Plan and carry out a directed Medline search that produces the articles to be discussed concerning the topic being discussed
 3. Determine whether the article(s) give us valid information concerning the question at hand.
 4. Determine whether the results are applicable to the patient in “your practice” case
5. Educational Strategies
- Case Based Learning
 - Interactive Workshops
 - The Creation of Learning Teams (groups of SENIOR RESIDENTS and INTERNS)
 - Train the trainer (Faculty training the senior house officers to teach the interns)
6. Implementation of the Curriculum
- Part I: Two didactic Sessions for interns: Question Building and Medline Searching
 - Part II: The Creation of Learning Teams:
 - 8 Teams (7 THIRD YEAR RESIDENTS, 7 Interns) who will work together for the entire year
 - 7 Workshop Sessions over the course of the year:
 - Each Workshop Topic has a faculty member as “Content Leader”
 - Therapy
 - Diagnosis
 - Evidence-Based Physical Exam
 - Prognosis
 - Harm
 - Systematic Review
 - Cost-Effectiveness Analysis
 - Content Leader’s Job: To meet with the 7 THIRD YEAR RESIDENTS who are scheduled to lead the Learning Teams and prepare them to teach that topic
 - Session Facilitators: 1 Faculty Facilitator for 2 Learning Teams. The job of the session facilitator is to problem solve during the sessions, to keep people running on time and to gather the two groups together for a 10-minute wrap up at the end of each session.
 - Part III: Modification of Current EBM Workshops during Ambulatory Block time to create co-teachers (faculty + SECOND YEAR RESIDENT). These sessions will continue to be very structured sessions with significant faculty input and direction.
7. Evaluation and Feedback
- Quality Improvement
 - Objective Measures of Learners skills, knowledge and attitudes
 - We will consider the development of case-based evaluations

How to choose critical appraisal worksheet

How do I know which critical appraisal sheet to use?

Answering the Clinical Question: Critical Appraisal- Survival Skills

A. Define the Clinical Question.

1. Patient, Population or Problem
2. Intervention, Prognostic Factor Exposure
3. Comparison Intervention (if appropriate)
4. Outcome you would like to measure or achieve
5. **Type of Question you are asking**
6. **Type of Study you would want to find**

What types of questions may we come up with?

(What Type of study would you want to find to answer that question?)

1. Clinical Examination	(Prospective cohort blind comparison to Gold Standard)
2. Diagnostic	Testing (Prospective cohort blind comparison to Gold Standard)
3. Prognosis	(Cohort Study>Case Control > Case Series)
4. Therapy	(RCT is really the only way we want to answer this question)
5. Etiology / Harm	(RCT> Cohort Study>Case Control>Case Series)
6. Prevention	(RCT> Cohort Study>Case Control>Case Series)
7. Cost	(Economic Analysis)
8. Self-Improve / Education	(RCT> Cohort Study)
9. Quality Improvement	(RCT> Cohort Study)
10. Health Services Research	(RCT> Cohort Study)
11. Differential Diagnosis	(Cohort Study)

Question to Consider:

**Was the type of study the strongest that could have been performed under the circumstances?
If not... Could you have designed the study better?**

Types of Studies:

Experimental Design:

Randomized Control Trial (RCT)

Guarantee Random distribution of factors known and unknown between groups aiming for equal distribution of factors between groups (remember that small studies may be random but not equal...)

This is an experimental method

Non- Experimental Design:

Cohort Study: follow one or more groups of individuals who have not yet suffered the adverse event and monitor the number of outcomes that occur over time. These need to be done when it is either not ethical or not practical to randomly assign patients to be "exposed" to something. Observational Design can be prospective or retrospective.

Case-Control Study: Collection of "cases" who have suffered the outcome and "controls" who have not. Investigators count the number of patients with a prognostic factor in the cases and the controls. These need to be done when the outcome of interest is rare or takes a long time to develop.

Case Series and Case Reports: Reports of patient scenarios that do not provide any comparison group.

B. Which critical appraisal sheet should you use for which study design?

Type of Sheet	Type of studies you would want to appraise with the sheet
Therapy	<ul style="list-style-type: none"> • Randomized Controlled Trial • Note this should be used for any “intervention” that has been tested by RCT including prevention, an RCT of a diagnostic test strategy, an RCT of a health services research intervention (e.g. change in clinic procedure), an RCT of an education intervention.
Diagnosis	<ul style="list-style-type: none"> • Prospective cohort, comparison to gold standard
Harm	<ul style="list-style-type: none"> • Case Control
Prognosis	<ul style="list-style-type: none"> • Cohort Study
Overview	<ul style="list-style-type: none"> • Systematic Review / Meta-analysis • Note: this is a summary methodology so you might have several different kinds of articles (Therapy RCT, Diagnosis cohort) BUT the critical appraisal is based on the way that the systematic review / meta-analysis was done.
Practice Guidelines	<ul style="list-style-type: none"> • Summary methodology based on a broad clinical topic instead of a focused clinical question • Note: Many different individual pieces of evidence will have contributed to the development of a practice guideline (often a Systematic Review of one or more of the key individual questions is associated) BUT the critical appraisal is based on the way the practice guideline was done.

C. Type of Cohort Study and “who” determines which group a participant is in?

Type of Sheet	Questions that will help you determine Validity of the Results
Observational Prospective Cohort to determine prognosis	<ul style="list-style-type: none"> • Impact of a Prognostic factor • Example: Individuals with ulcerative colitis and the development of colon cancer • Who decides? Fate.
Observational Prospective Cohort to determine prognosis	<ul style="list-style-type: none"> • Impact of an exposure • Example: Smokers and risk of lung cancer • Who decides? The patient/ person him or herself
Interventional Prospective Cohort to determine effect of an intervention on prognosis	<ul style="list-style-type: none"> • Impact of an intervention • Example: Steroid inhaler for asthma • Who decides? Random process (if Randomization is done correctly)

How to use the rational clinical examination education guides

It would be better if you began to teach others only after you yourself have learned something.

—Albert Einstein to Arthur Cohen, age 12, who submitted a paper to Einsteinⁱ

Teachers as learners; learners as teachers

Take a moment to recall teachers who truly influenced your understanding. What about those individuals made them great teachers? Did they simplify key concepts? Did they help you understand why you needed to know or connect ideas to show a common thread? Did they make it fun and interactive? Likely, it was a combination of these factors.

Successful teaching is not a purely spontaneous event, although great teachers will make it seem that way. Rather, effective teaching is deliberate: it follows from practice, patience, and planning. Thus, we created the Education Guides for the Rational Clinical Examination with a systematic approach that provides teachers with tools, tips, and ideas for making the contents of the book real, meaningful, and exciting to their learners.

We intentionally sought learners as collaborators for producing the Education Guides; approximately 90% of the chapters involved Duke University Department of Medicine residents or fellows, often as the lead author of the teaching materials. Learners' active involvement kept the Education Guides relevant to clinical trainees and generalist physicians. It also served to blur the lines between teachers and learners. To teach the material, the Education Guides authors first needed to learn the material! After the authors created the Education Guides, the Editors reviewed the slides for education content, flow and relevance. Finally, authors of the original Rational Clinical Examination article or its Update reviewed the Guide to assure that the content and emphasis were consistent with their prior work.

Both teachers and learners who are in a hurry may access the chapter content through the Education Guides. However, the Guides only complement, not replace the chapters of the Rational Clinical Examination. Readers or educators who choose to use the slides independently from the book will not be well-prepared, as they will miss some of the salient features.

Teaching Tip #1

Be familiar with core content. Preparing to teach is first a learning exercise. The teaching materials provide a summary of the key content to complement each chapter of the book.

Educators and learners are encouraged to spend time understanding the chapters that are particularly relevant to their everyday practice and teaching. Most educators can easily identify clinical conditions or findings that they repeatedly encounter. For example, an attending covering the inpatient hospital service may first become familiar with the chapter on deep vein thrombosis while a resident or medical student preparing for a clinical rotation in their emergency room might study in advance the chapter on acute dyspnea.

Teaching Tip #2

Prioritize your reading and learning to focus on clinical syndromes and settings you most frequently encounter. Identify topics that are predictably present in your clinical education environment and become familiar with the prior probabilities and likelihood ratios that apply.

What's in the education guides?

All the information summarized in the Education Guides comes directly from the original Rational Clinical Examination article or its Update. Each set of teaching materials begins with one or more clinical case scenarios and a series of questions to pose to learners. We always ask that learners explicitly state their impression of likelihoods or probabilities of the target conditions. This immediately forces teachers

into an environment of active engagement. Our interactive approach at the beginning of each Education Guide produces a skill building exercise whereby 1) learners think in terms of probability, 2) we promote retention and understanding by getting learners to name their educated guess and check this estimate against the evidence, and 3) we ultimately either reinforce or redirect their prior assumptions. Each set of teaching materials ends with the resolution of the clinical scenarios followed by “take home messages” and a “bottom line” for the chapter.

In addition to the clinical content and data in the Education Guides, we added teachers' notes and tips. Microsoft® Office PowerPoint® has a feature that allows each slide to be viewed with a Notes page. We used the Notes pages to identify basic principles, potential obstacles, and strategies for interactive teaching such as the use of slide animation. The Notes pages also provide teachers with an additional layer of information to enhance or further explain the bullet points or tables on the slides to assist in preparation for a teaching session.

Field testing of the Education Guides

The teaching strategies and stumbling blocks described as part of the Education Guides have been field tested for clarity and relevance among Duke Internal Medicine residents at the Durham Veterans Affairs Medical Center. The general interactive teaching strategies were developed and tested over the past 5-10 years by experts participating in McMaster University and Duke University workshops on teaching evidence-based practice. Although some of the teaching strategies have been published as a part of the evidence-based teaching tips projectⁱⁱ the strategies have not undergone formal testing and thus reflect expert opinion.

Planning for delivery: Maximizing interactivity in classroom settings

The Education Guides use the primary format of PowerPoint® slides. However, educators are encouraged to avoid a purely didactic lecture style for this content (or any content, for that matter!) In fact, if the guides are only used for didactic presentations, we will have failed in our attempt to encourage strategies to address learner engagement. The teaching tips focus on two key elements of engagement: relevance and interactivity.

We systematically designed the case scenarios by including clinical elements that highlight key points in each chapter. When more than one case is used, the cases compare and contrast different aspects of clinical decision making. For example, a chapter may include cases that reflect examples of low, intermediate and high prior probability of disease. This allows the educator to illustrate the impact of differing prior probability on posttest probability. Similarly, the cases might reflect differing patient characteristics that require consideration of action thresholds for pursuing additional tests or implementing a treatment strategy.

Teaching Tip #3

Focus on relevance using a case-based format. Clinical examination is a skill that should be taught in context. In classroom settings, anchor your teaching with the clinical cases provided in the materials, or cases of relevance to you and your learners.

Educators are encouraged to view the PowerPoint® slides as part of a preparatory toolkit, rather than 'readymade' slides set for presentation. In fact, educators may most effectively use the materials for teaching without actually projecting a single slide. For example, a very effective classroom teaching session might involve describing the 3 cases that are used for the chapter on community-acquired pneumonia in adults. The learners could be broken up into 3 small groups, each assigned to discuss one of the patient cases. As a first step, the learners could be asked to discuss the cases without any further information and to estimate the probability that each patient has community-acquired pneumonia.

These estimates can be written on a flip chart and discussion can take place about what elements went into the decision making for each group.

Teaching Tip #4

Ask learners to commit to probabilities. Creating a safe learning environment in which learners can discuss their initial assessments is important to help them build on their base knowledge in each session.

The educator can then discuss the concept of the likelihood ratio, the prior probability of disease and the individual likelihood ratios for the clinical examination items. The educator should ask whether this information would alter the learners' assessment of the likelihood of disease.

Teaching Tip #5

Focus on learner interaction, minimizing or eliminating didactic teaching. The teaching tools should serve as substrate interactive teaching. Educators can combine some didactic teaching for emphasis, orientation and reinforcement of principles, but primary strategies should be interactive.

In the example of community-acquired pneumonia, the likelihood ratios for the individual findings are not very useful so this creates an opportunity for discussing multivariate analyzes and clinical prediction rules. The learners could be given summaries of the Diehr multivariate model and the Heckerling clinical prediction model and break into their 3 groups to repeat their discussion on the assessment of the probability of pneumonia.

Teaching Tip #6

Focus the clinical examination on useful items while pointing out findings that may not be helpful. As learners familiarize themselves with likelihood ratios, educators should identify clinical examination that impacts those assessments and also dispel myths about examination items that don't.

Using the Diehr multivariate model as an example, the educator could have each group come up with the likelihood ratio to apply to their patient. The educator can then hand out a blank nomogram such as the one that is included in the PRIMER (A Primer on the Precision and Accuracy of the Clinical Examination: Introduction) and have the trainees plot the results for each case. The nomogram serves as a visual tool to illustrate the concept that a likelihood ratio, applied to a prior probability, generates the posttest probability of disease.

Teaching Tip #7

Use the nomograms to illustrate movement from pretest to posttest probabilities. The nomogram can be a valuable visual and conceptual tool when working through individual patient cases.

Planning for delivery: Finding ways to practice the skill of taking history and physical examination

For optimal professional development, trainees require orientation to key concepts of clinical examination, skills practice and feedback from their faculty mentors and role models. Some of the topics are particularly suited for classroom practice of clinical exams, such as examination of the shoulder or knee. To facilitate these learning exercises, the Rational Clinical Examination and Education Guides have pictures and illustrations that highlight the technical points of clinical maneuvers.

Some of the clinical examination items will require teaching directly in the context of patient care, for example, learning to assess central venous pressure or ascites. Patient-centered teaching can be complemented by bringing the evidence from the Education Guides to a teaching session or rounds either proceeding or following a trip to the bedside. Educational assignments written on a prescription pad, called education prescriptions, encourage the learner to follow up on a finding identified during

ward rounds or clinic sessions. The prescription should note the clinical question and suggest the relevant Rational Clinical Examination articles. Just as in clinical medicine where the physician follow-up on the treatment response, the education prescriber should follow-up with the learner at their next clinical session.

Teaching Tip #8

Find ways to practice hands-on maneuvers with your learners in both the classroom and clinical settings. Learners need to practice and receive feedback on clinical examination. When patients are respectfully included, skills can be refined in the clinical environment as well as in the classroom setting.

A barrier when trying to incorporate evidence into clinical teaching is discomfort with statistical principles and the frequent misperception that evidence-based practice is equivalent to statistics. Throughout the Education Guides, we included descriptions of common statistical concepts that the teachers and learners will encounter. The PRIMER includes an entire set of descriptions and teaching strategies that can serve to assist educators in confronting these principles.

However, we emphasize that while understanding the statistical concepts is helpful, it should not take away from the clinical application and focus of a teaching session. The Education Guides should help educators and learners become better users of the medical literature on the clinical examination, rather than to become statisticians or researchers.

Teaching Tip #9

Avoid statistical jargon. The goal is to assist learners to become effective at incorporation of evidence into clinical practice.

More than any other goal in the creation of the Education Guides for the Rational Clinical Examination, we hope that educators and learners will have fun with the book and the tools it supplies. The Rational Clinical Examination series provides a plethora of teaching opportunities that uniquely combine evidence and the medical literature with direct patient care. Enjoy yourself and good things will follow.

Teaching Tip #10

Have fun. Strive to employ new and creative ways to engage your learners, involve them in the excitement of clinical decision making and the fun of lifelong learning.

ⁱ Calaprice, Alice. The new quotable Einstein. Princeton, NJ: Princeton University Press; 2005, p. 66.

ⁱⁱ Wyer PC, Keitz SA, Hatala R, Hayward R, Barratt A, Montori V, Wooltorton E, Guyatt G. Tips for learning and teaching evidence-based medicine: introduction to the series. CMAJ. 2004;171(4):347-348.

Educational Prescription: Duke University Medical Center

What is an educational prescription?

It's a learning assignment co-written by preceptor and learner that

- specifies the clinical problem that generated the question.
- states the question, in all 4 of its key elements (PICO).
- specifies who is responsible for answering it.
- reminds everyone of the deadlines for answering it (taking into account the urgency of the clinical problem that generated it).

Why use educational prescriptions?

Questions arise but they don't always get followed up because clinical constraints and fatigue often limit our opportunities. Using educational prescriptions helps us keep track of our questions so that we can answer them when an opportunity develops. Prescriptions help learners practice the important lifelong habit of using EBM on a daily basis to help answer clinical questions.

Tips for using educational prescriptions

- Include them as a regular part of rounds, sign-outs, and supervision.
- Ask your learners to write educational prescriptions for you.
- Keep a copy of the prescription for you and the service's chief resident.
- Use the opportunity to introduce the learner to a University librarian who can help 'fill' the prescription.
- Follow-up with the learner on the pre-specified date.

Resources: <https://guides.mclibrary.duke.edu/ebm/home>

(Modified from <http://www.cebm.utoronto.ca/practise/formulate/eduprescript.htm>)

Patient's Initials/MRN: _____ Learner: _____

Clinical Question

Patient or Problem:
Intervention:
Comparison:
Outcome(s):
Type of question:
Study type:

Date and place to present findings: _____

Presentation will cover:

1. search strategy
2. search results
3. appraisal of the validity of the evidence
4. appraisal of the importance of the results
5. application to the patient/problem
6. your self-evaluation of this process

First copy to learner; second copy to attending; third copy to chief resident

Appendix

Critical Appraisal Forms

Clinical Decision Analysis

Citation:

Is this a newly derived instrument? (Level IV)	
Was validation restricted to the retrospective use of statistical techniques on the original database? (If so this is a Level IV rule). If so, consider the following standards for initial development of a decision rule.	
Were all important predictors included in the derivation process?	
Were all important predictors present in significant proportion of the study population?	
Does the rule make clinical sense?	
Has the instrument been validated? (Level II or III) If so, consider the following:	
Did validation include prospective studies on several different populations from that used to derive it (II), or was it restricted to a single population (III)?	
How well did the validation exercise meet the following criteria? Were the patients chosen in an unbiased fashion and do they represent a wide spectrum of severity of disease?	
Was there a blinded assessment of the criterion standard or outcome event (or was the outcome all-cause mortality) for all patients?	
Was there an explicit and accurate interpretation of the predictor variables and the actual rule without knowledge of the outcome?	
Was there 100% follow-up of those enrolled?	
How powerful is the rule (in terms of sensitivity and specificity; likelihood ratios; proportions with alternative outcomes; or relative risks or absolute outcome rates)?	
Has an impact analysis demonstrated change in clinical behavior or patient outcomes as a result of using the instrument? (Level I)	
How well did the study guard against bias in terms of differences at the start (concealed randomization, adjustment in analysis) or as the study proceeded (blinding, cointervention, loss to follow-up)	
What was the impact on clinician behaviour and patient-important outcomes?	

Adapted from McMaster Evidence-based Clinical Practice Workshops and Users' Guide to the Medical Literature 3rd Ed.

Clinical Practice Guidelines

Citation:

Is the clinical question clear and comprehensive?	
Is the recommended intervention clear and actionable?	
Is the alternative clear?	
Were all of the relevant outcomes important to patients explicitly considered?	
Was the recommendation based on the best current evidence?	
Are values and preferences associated with the outcomes appropriately specified?	
Do the authors indicate the strength of their recommendations?	
Is the evidence supporting the recommendation easily understood?	
For strong recommendations, is the strength appropriate?	
For weak recommendations, does the information provided facilitate shared decision making?	
Was the influence of the conflict of interests minimized?	

Adapted from McMaster Evidence-based Clinical Practice Workshops and Users' Guide to the Medical Literature 3rd Ed.

Diagnostic Test

Citation:

How serious is the risk of bias?	
Did participating patients constitute a representative sample of those presenting with a diagnostic dilemma?	
Did investigators compare the test to an appropriate, independent reference standard?	
Were those interpreting the test and reference standard blind to the other results?	
Did all patients receive the same reference standard irrespective of the results of the test results?	
What are the results?	
What likelihood ratios were associated with the range of possible test results?	
How can I apply the results to patient care?	
Will the reproducibility of the test results and its interpretation be satisfactory in my clinical setting?	
Are the study results applicable to patients in my practice?	
Will the test results change my management strategy?	
Will patients be better off as a result of the test?	

Adapted from McMaster Evidence-based Clinical Practice Workshops and Users' Guide to the Medical Literature 3rd Ed.

Differential Diagnosis

Citation:

How serious is the risk of bias?	
Did the study patients represent the full spectrum of those with this clinical problem?	
Was the diagnostic evaluation definitive?	
What are the Results?	
What were the diagnoses and their probabilities?	
How precise are the estimates of disease probability?	
How can I apply the results to patient care?	
Are the study patients and clinical setting similar to mine?	
Is it unlikely that the disease possibilities or probabilities have changed since this evidence was gathered?	

Adapted from McMaster Evidence-based Clinical Practice Workshops and Users' Guide to the Medical Literature 3rd Ed.

Are the results valid?	
Did the recommendations consider all relevant patient groups, management options, and possible outcomes?	
Did the investigators adopt a sufficiently broad viewpoint?	
Are results reported separately for relevant patient subgroups?	
Is there a systematic review and summary of evidence linking options to outcomes for each relevant question?	
Were costs measured accurately?	
Did investigators consider the timing of costs and outcomes?	
What are the results?	
What were the incremental costs and effects of each strategy?	
Do incremental costs and effects differ between subgroups?	
How much does allowance for uncertainty change the results?	
How can I apply the results to patient care?	
Are the treatment benefits worth the harms and the costs?	
Can I expect similar costs in my setting?	

Adapted from McMaster Evidence-based Clinical Practice Workshops and the Users' Guide to the Medical Literature 3rd Ed.

How serious is the risk of bias?	
Aside from the exposure of interest, did the exposed and control groups start and finish with the same risk for the outcome?	
Were the patients similar for prognostic factors that are known to be associated with the outcome (or did statistical adjustment level the playing field)?	
Were the circumstances and methods for detecting the outcome similar?	
Was the follow-up sufficiently complete?	
What are the Results?	
How strong is the association between exposure and outcome?	
How precise is the estimate of risk?	
How can I apply the results to my patient care?	
Were the study patients similar to patients in my practice?	
Was follow-up sufficiently long?	
Is the exposure similar to what might occur in my patient?	
What is the magnitude of the risk?	
Are there any benefits that are known to be associated with the exposure?	

Adapted from McMaster Evidence-based Clinical Practice Workshops and Users' Guide to the Medical Literature 3rd Ed.

Harm, Case-Control Study

Citation:

How serious is the risk of bias?	
Did the cases and control group have the same risk (chance) for being exposed in the past?	
Were cases and controls similar with respect to the indication or circumstances that would lead to exposure?	
Were the circumstances and methods for determining exposure similar for cases and controls?	
What are the Results?	
How strong is the association between exposure and outcome?	
How precise is the estimate of risk?	
How can I apply the results to my patient care?	
Were the study patients similar to patients in my practice?	
Was follow-up sufficiently long?	
Is the exposure similar to what might occur in my patient?	
What is the magnitude of the risk?	
Are there benefits that offset the risks of the exposure?	

Adapted from McMaster Evidence-based Clinical Practice Workshops and Users' Guide to the Medical Literature 3rd Ed.

Prognosis

Citation:

How serious is the risk of bias?	
Was the sample of patients' representative?	
Were patients classified into prognostically homogeneous groups?	
Was follow-up sufficiently complete?	
Were outcome criteria objective and unbiased?	
What are the results?	
How likely are the outcomes over time?	
How precise are the estimates of likelihood?	
How can I apply the results to patient care?	
Were the study patients and their management similar to those in my practice?	
Was the follow-up sufficiently long?	
Can I use the results in the management patients in my practice?	

Adapted from McMaster Evidence-based Clinical Practice Workshops and Users' Guide to the Medical Literature 3rd Ed.

Qualitative Methods

Citation:

Is the qualitative research relevant?	
Are the results credible?	
Is there a specific qualitative research method cited?	
Was the choice of participants or observations explicit and comprehensive?	
Were research ethics approval obtained?	
Was data collection sufficiently comprehensive and detailed?	
Were the data analyzed appropriately and the findings corroborated adequately?	
What are the results?	
What are the results?	
How can I apply the results to patient care?	
How does the study offer helpful theory?	
Does the study help me to understand the context of my practice?	
Does the study help me to understand social phenomena in my practice?	

Adapted from McMaster Evidence-based Clinical Practice Workshops and Users' Guide to the Medical Literature 3rd Ed.

Screening

Citation:

How serious is the risk of bias?	
Is there randomized controlled trial evidence that the intervention benefits people with asymptomatic disease?	
What are the recommendations, and will they help you in caring for patients?	
Were the data identified, selected, and combined in an unbiased fashion?	
What are the benefits?	
What are the harms?	
How do benefits and harms compare in different people and with different screening strategies?	
What is the effect of individuals' values and preferences?	
What is the effect of uncertainty associated with the evidence?	
What is the cost-effectiveness?	

Adapted from McMaster Evidence-based Clinical Practice Workshops and Users' Guide to the Medical Literature 3rd Ed.

Systematic Review

Citation:

Assessing the Credibility of the Systematic Review Process	
Did the review address a focused clinical question (i.e. can be framed in PICO format)?	
Was the search for relevant studies detailed and exhaustive?	
Were selection and assessment of studies reproducible?	
Was the risk of bias of the primary studies assessed?	
Did the review address possible explanations of between-study differences in results using prespecified hypotheses?	
Did the review describe a process to assess confidence in effect estimates? (e.g. GRADE tool to assess quality of the body of evidence)	
Understanding the Summary Estimate of a Meta-analysis	
What is the magnitude of treatment effect? (what is the pooled estimate?)	
How precise are the results? (i.e. confidence interval around the pooled estimate)	
Rating Confidence in the Estimates (the Quality of a Body of Evidence)	
How serious is the risk of bias in the body of evidence?	
Are the results consistent across studies? (i.e. heterogeneity or inconsistency)	
Do the results directly apply to my patient? (i.e. PICO, generalizability, indirectness)	
Is there a concern about reporting or publication bias?	
Are there reasons to increase or decrease the confidence of the rating? (<i>Randomized trials start high and observational studies start low</i>)	
Overall, what is the quality of the body of evidence by outcome? (High, moderate, low, very low)	
How can I apply the results to my patient care?	
Did the review present results that are ready for clinical application? (e.g. patient important outcomes, absolute benefit /risk)	
Are the study patients similar to my patient and are likely benefits worth potential harms/costs?	

Adapted from McMaster Evidence-based Clinical Practice Workshops and the Users' Guide to the Medical Literature 3rd Ed.

Therapy

Citation:

How serious is the risk of bias?	
Did intervention and control groups start with the same prognosis?	
Were patients randomized?	
Was randomization concealed?	
Were patients in the study groups similar at baseline with respect to prognostic factors?	
Was prognostic balance maintained as the study progressed?	
To what extent was the study blinded?	
Were groups prognostically balanced at the study's conclusion?	
Was follow-up complete?	
Were patients analyzed in the groups to which they were randomized?	
Was the trial stopped early?	
What are the results?	
How large was the treatment effect?	
How precise was the treatment effect?	
How can I apply the results to my patient care?	
Were the study patients similar to my patient?	
Were all patient-important outcomes considered?	
Are the likely benefits worth the potential harms and costs?	

Adapted from McMaster Evidence-based Clinical Practice Workshops and the Users' Guide to the Medical Literature 3rd Ed.

Therapy, Non-inferiority trials

Citation:

How serious is the risk of bias?	
Did intervention and control groups begin the study with a similar prognosis?	
Were patients randomized?	
Was randomization concealed?	
Were patients similar at baseline with respect to known prognostic factors?	
Was prognostic balance maintained as the study progressed?	
Were patients, caregivers, collectors of outcome data, adjudicators of outcome, and data analysts aware of group allocation?	
Were groups prognostically balanced at the study's conclusion?	
Was follow-up complete?	
Was the trial stopped early for benefit?	
Were patients analyzed in the groups to which they were randomized?	
Did the investigators guard against an unwarranted conclusion of non-inferiority?	
Was the effect of the standard treatment preserved?	
Did the investigators analyze patients according to the treatment they received, as well as to the groups to which they were assigned?	

What are the results?	
How large was the treatment effect?	
How precise was the estimate of the treatment effect?	
How can I apply the results to my patient care?	
Were the study patients similar to my patient?	
Were all patient-important outcomes considered?	
Are the likely advantages of the novel treatment worth the potential harms and costs?	

Adapted from McMaster Evidence-based Clinical Practice Workshops and the Users' Guide to the Medical Literature 3rd Ed.

Teaching Papers

Diagnosis

(1990). "Value of the ventilation/perfusion scan in acute pulmonary embolism. Results of the prospective investigation of pulmonary embolism diagnosis (PIOPED). The PIOPED Investigators." *JAMA* 263(20): 2753-2759.

To determine the sensitivities and specificities of ventilation/perfusion lung scans for acute pulmonary embolism, a random sample of 933 of 1493 patients was studied prospectively. Nine hundred thirty-one underwent scintigraphy and 755 underwent pulmonary angiography; 251 (33%) of 755 demonstrated pulmonary embolism. Almost all patients with pulmonary embolism had abnormal scans of high, intermediate, or low probability, but so did most without pulmonary embolism (sensitivity, 98%; specificity, 10%). Of 116 patients with high-probability scans and definitive angiograms, 102 (88%) had pulmonary embolism, but only a minority with pulmonary embolism had high-probability scans (sensitivity, 41%; specificity, 97%). Of 322 with intermediate-probability scans and definitive angiograms, 105 (33%) had pulmonary embolism. Follow-up and angiography together suggest pulmonary embolism occurred among 12% of patients with low-probability scans. Clinical assessment combined with the ventilation/perfusion scan established the diagnosis or exclusion of pulmonary embolism only for a minority of patients--those with clear and concordant clinical and ventilation/perfusion scan findings.

Question: Diagnosis

Study Type: Prospective Cohort

Learner Level: Beginner / Intermediate; Classic paper

Notes: Classic Diagnosis paper for teaching likelihood ratios. This is a paper of very strong methodology with great results that lend themselves to perfect discussion of the value of likelihood ratios. Although the paper is 13 years old, there are few better for discussing LRs. In addition, the results of this trial continue to impact our thinking and the reading of the medical literature when looking at other papers about diagnostic tests for evaluation of PE (e.g. Spiral CT.); One of our all-time favorite teaching papers

Aberle, D. R., et al. (2011). "Reduced lung-cancer mortality with low-dose computed tomographic screening." *N Engl J Med* 365(5): 395-409.

BACKGROUND: The aggressive and heterogeneous nature of lung cancer has thwarted efforts to reduce mortality from this cancer through the use of screening. The advent of low-dose helical computed tomography (CT) altered the landscape of lung-cancer screening, with studies indicating that low-dose CT detects many tumors at early stages. The National Lung Screening Trial (NLST) was conducted to determine whether screening with low-dose CT could reduce mortality from lung cancer. **METHODS:** From August 2002 through April 2004, we enrolled 53,454 persons at high risk for lung cancer at 33 U.S. medical centers. Participants were randomly assigned to undergo three annual screenings with either low-dose CT (26,722 participants) or single-view posteroanterior chest radiography (26,732). Data were collected on cases of lung cancer and deaths from lung cancer that occurred through December 31, 2009. **RESULTS:** The rate of adherence to screening was more than 90%. The rate of positive screening tests was 24.2% with low-dose CT and 6.9% with radiography over all three rounds. A total of 96.4% of the positive screening results in the low-dose CT group and 94.5% in the radiography group were false positive results. The incidence of lung cancer was 645 cases per 100,000 person-years (1060 cancers) in the low-dose CT group, as compared with 572 cases per 100,000 person-years (941 cancers) in the radiography group (rate ratio, 1.13; 95% confidence interval [CI], 1.03 to 1.23). There were 247 deaths from lung cancer per 100,000 person-years in the low-dose CT group and 309 deaths per 100,000 person-years in the radiography group, representing a relative reduction in mortality from lung cancer with low-dose CT screening of 20.0% (95% CI, 6.8 to 26.7; P=0.004). The rate of death from any cause was reduced in the low-dose CT group, as compared with the radiography group, by 6.7% (95% CI, 1.2 to 13.6; P=0.02). **CONCLUSIONS:** Screening with the use of low-dose CT reduces mortality from lung cancer. (Funded by the National Cancer Institute; National Lung Screening Trial ClinicalTrials.gov number, NCT00047385.).

Question: Screening

Study Type: RCT

Learner Level: Intermediate

Notes: great to contrast cohort (Henschke) vs RCT data for screening

Acebedo-Urdiales, M. S., et al. (2014). "Practical knowledge of experienced nurses in critical care: a qualitative study of their narratives." *BMC Medical Education* 14: 173.

BACKGROUND: Scholars of nursing practices have claimed practical knowledge is source of knowledge in its own right, nevertheless we know little about this knowledge associated with day-to-day practice. The purpose of this study is to describe knowledge that the more experienced nurses in the ICU make use of and discover the components of care it includes. Understanding this knowledge can contribute to improving the working practices of nurses with less experience. **METHODS:** We used a phenomenologic and hermeneutic approach to conduct a qualitative study. Open in-depth dialogue interviews were conducted with 13 experienced ICU nurses selected by intentional sampling. Data was compiled on significant stories of their practice. The data analysis enabled units of meaning to be categorized and grouped into topics regarding everyday practical knowledge. **RESULTS:** Knowledge related to everyday practice was evaluated and grouped into seven topics corresponding to how the ICU nurses understand their patient care:

1) Connecting with, calming and situating patients who cannot communicate; 2) Situating and providing relief to patients in transitions of mechanical respiration and non-invasive ventilation; 3) Providing reassurance and guaranteeing the safety of immobilised patients; 4) The "connection" with patients in comas; 5) Taking care of the body; 6) The transition from saving life to palliative care; and 7) How to protect and defend the patient from errors. The components of caretaking that guarantee success include: the calm, care and affection with which they do things; the time devoted to understanding, situating and comforting patients and families; and the commitment they take on with new staff and doctors for the benefit of the patient. **CONCLUSIONS:** These results show that stories of experiences describe a contextual practical knowledge that the more experienced nurses develop as a natural and spontaneous response. In critical patients the application of everyday practical knowledge greatly influences their well-being. In those cases in which the nurses describe how they have protected the patients from error, this practical knowledge can mean the difference between life and death. The study highlights the need to manage practical knowledge and undertake further research. The study is useful in keeping clinical practice up-to-date.

Question: .

Study Type: Qualitative

Learner Level: .

Notes: Qualitative Study

Andriole, G. L., et al. (2009). "Mortality results from a randomized prostate-cancer screening trial." *N Engl J Med* 360(13): 1310-1319.

BACKGROUND: The effect of screening with prostate-specific-antigen (PSA) testing and digital rectal examination on the rate of death from prostate cancer is unknown. This is the first report from the Prostate, Lung, Colorectal, and Ovarian (PLCO) Cancer Screening Trial on prostate-cancer mortality. **METHODS:** From 1993 through 2001, we randomly assigned 76,693 men at 10 U.S. study centers to receive either annual screening (38,343 subjects) or usual care as the control (38,350 subjects). Men in the screening group were offered annual PSA testing for 6 years and digital rectal examination for 4 years. The subjects and health care providers received the results and decided on the type of follow-up evaluation. Usual care sometimes included screening, as some organizations have recommended. The numbers of all cancers and deaths and causes of death were ascertained. **RESULTS:** In the screening group, rates of compliance were 85% for PSA testing and 86% for digital rectal examination. Rates of screening in the control group increased from 40% in the first year to 52% in the sixth year for PSA testing and ranged from 41 to 46% for digital rectal examination. After 7 years of follow-up, the incidence of prostate cancer per 10,000 person-years was 116 (2820 cancers) in the screening group and 95 (2322 cancers) in the control group (rate ratio, 1.22; 95% confidence interval [CI], 1.16 to 1.29). The incidence of death per 10,000 person-years was 2.0 (50 deaths) in the screening group and 1.7 (44 deaths) in the control group (rate ratio, 1.13; 95% CI, 0.75 to 1.70). The data at 10 years were 67% complete and consistent with these overall findings. **CONCLUSIONS:** After 7 to 10 years of follow-up, the rate of death from prostate cancer was very low and did not differ significantly between the two study groups. (ClinicalTrials.gov number, NCT00002540.)

Question: Screening

Study Type: RCT

Learner Level: Advanced

Notes: Good study to discuss challenges in the design and execution of screening studies i.e. statistical power considerations, contamination of control arm. Best taught in conjunction with: Schroder, F. H., Hugosson, J., Roobol, M. J. et al.: Screening and prostate-cancer mortality in a randomized European study. *N Engl J Med*, 360: 1320, 2009

Aronson, P. L., et al. (2018). "Risk Stratification of Febrile Infants \leq 60 Days Old Without Routine Lumbar Puncture." *Pediatrics* 142(6).

OBJECTIVES: To evaluate the Rochester and modified Philadelphia criteria for the risk stratification of febrile infants with invasive bacterial infection (IBI) who do not appear ill without routine cerebrospinal fluid (CSF) testing. **METHODS:** We performed a case-control study of febrile infants \leq 60 days old presenting to 1 of 9 emergency departments from 2011 to 2016. For each infant with IBI (defined as a blood [bacteremia] and/or CSF [bacterial meningitis] culture with growth of a pathogen), controls without IBI were matched by site and date of visit. Infants were excluded if they appeared ill or had a complex chronic condition or if data for any component of the Rochester or modified Philadelphia criteria were missing. **RESULTS:** Overall, 135 infants with IBI (118 [87.4%] with bacteremia without meningitis and 17 [12.6%] with bacterial meningitis) and 249 controls were included. The sensitivity of the modified Philadelphia criteria was higher than that of the Rochester criteria (91.9% vs 81.5%; $P = .01$), but the specificity was lower (34.5% vs 59.8%; $P < .001$). Among 67 infants >28 days old with IBI, the sensitivity of both criteria was 83.6%; none of the 11 low-risk infants had bacterial meningitis. Of 68 infants ≤ 28 days old with IBI, 14 (20.6%) were low risk per the Rochester criteria, and 2 had meningitis. **CONCLUSIONS:** The modified Philadelphia criteria had high sensitivity for IBI without routine CSF testing, and all infants >28 days old with bacterial meningitis were classified as high risk. Because some infants with bacteremia were classified as low risk, infants discharged from the emergency department without CSF testing require close follow-up.

Question: Diagnosis - risk stratification

Study Type: Case-Control Study

Learner Level: Intermediate

Notes:

Ashdown, H. F., et al. (2012). "Pain over speed bumps in diagnosis of acute appendicitis: diagnostic accuracy study." *BMJ* 345: e8012.

OBJECTIVE: To assess the diagnostic accuracy of pain on travelling over speed bumps for the diagnosis of acute appendicitis. **DESIGN:** Prospective questionnaire based diagnostic accuracy study. **SETTING:** Secondary care surgical assessment unit at a district general hospital in the UK. **PARTICIPANTS:** 101 patients aged 17-76 years referred to the on-call surgical team for assessment of possible appendicitis. **MAIN OUTCOME MEASURES:** Sensitivity, specificity, positive and negative predictive values, and positive and negative likelihood ratios for pain over speed bumps in diagnosing appendicitis, with histological diagnosis of appendicitis as the reference standard. **RESULTS:** The analysis included 64 participants who had travelled over speed bumps on their journey to hospital. Of these, 34 had a confirmed histological diagnosis of appendicitis, 33 of whom reported increased pain over speed bumps. The sensitivity was 97% (95% confidence interval 85% to 100%), and the specificity was 30% (15% to 49%). The positive predictive value was 61% (47% to 74%), and the negative predictive value was 90% (56% to 100%). The likelihood ratios were 1.4 (1.1 to 1.8) for a positive test result and 0.1 (0.0 to 0.7) for a negative result. Speed bumps had a better sensitivity and negative likelihood ratio than did other clinical features assessed, including migration of pain and rebound tenderness. **CONCLUSIONS:** Presence of pain while travelling over speed bumps was associated with an increased likelihood of acute appendicitis. As a diagnostic variable, it compared favourably with other features commonly used in clinical assessment. Asking about speed bumps may contribute to clinical assessment and could be useful in telephone assessment of patients.

Question: Diagnosis

Study Type: .

Learner Level: .

Notes: This study is good methodologically, and allows discussing the issue of using a combined gold standard.

Azoulay, E., et al. (2009). "Polymerase chain reaction for diagnosing pneumocystis pneumonia in non-HIV immunocompromised patients with pulmonary infiltrates." *Chest* 135(3): 655-661.

RATIONALE: Pneumocystis jiroveci polymerase chain reaction (PCR) has higher sensitivity than conventional stains but cannot distinguish colonization from infection. **METHODS:** We compared P jiroveci PCR and conventional stains in HIV-uninfected immunocompromised patients. **RESULTS:** Among the 448 patients, 296 (66%) patients had hematologic malignancies; 72 (16.1%), bone marrow transplants; 44 (9.8%), solid tumors; 21 (4.7%), renal transplants; and 15 (3.4%) were taking immunosuppressants for systemic diseases. Diagnostic strategy consisted of BAL in 351 patients and induced sputum (IS) in 97 patients. Conventional pneumocystic pneumonia (PCP) stain was positive in 39 (8.7%) patients, including 34 with positive PCR. In addition, PCR was positive in 32 patients, including 21 with complete follow-up, of whom 14 were diagnosed with probable or definitive PCP (a 36% increase). PCR was 87.2% sensitive and 92.2% specific; positive and negative predictive values were 51.5% and 98.7%, respectively. Sensitivity and negative predictive value were 100% on IS. **CONCLUSIONS:** In HIV-uninfected immunocompromised patients with acute pulmonary infiltrates, P jiroveci PCR correlates with clinical evidence of PCP. A negative PCR allows withdrawing anti-PCP therapy.

Question: Diagnosis

Study Type: Prospective Cohort Study

Learner Level: Beginner

Notes: Good for teaching diagnosis Math and to show the relationships between sensitivity, specificity, positive negative predictive values and likelihood ratio. (Tip: make a 2x2 table from Figure 1 which allows you to do the math). Also good for discussion about the importance of pretest probability.

Baarslag, H. J., et al. (2002). "Prospective study of color duplex ultrasonography compared with contrast venography in patients suspected of having deep venous thrombosis of the upper extremities." *Ann Intern Med* 136(12): 865-872.

BACKGROUND: The optimal strategy for diagnosis of deep venous thrombosis (DVT) is less well established for the upper extremities than for the lower extremities. Duplex color ultrasonography can be difficult to perform in the upper extremities because of their anatomy, and contrast venography is often indicated. Moreover, limited data exist on the use of duplex color ultrasonography in this setting. **OBJECTIVE:** To determine the accuracy of duplex ultrasonography for diagnosis of DVT of the upper extremities. **DESIGN:** Prospective study of duplex ultrasonography compared with venography. **SETTING:** A teaching hospital in Amsterdam, the Netherlands. **PATIENTS:** 126 consecutive inpatients and outpatients with suspected DVT of the upper extremities. **MEASUREMENTS:** Contrast venography was obtained after duplex ultrasonography and was judged independently. A three-step protocol, involving compression ultrasonography, color ultrasonography, and color Doppler ultrasonography, was used. Sensitivity, specificity, and likelihood ratios for ultrasonography as a whole were calculated. The independent value of each step was assessed. **RESULTS:** Venography and ultrasonography were not feasible in 23 of 126 patients (18%) and 1 of 126 patients (0.8%), respectively. Results of ultrasonography were inconclusive in 3 patients. Venography demonstrated thrombosis in 44 of 99 patients (44%); in 36 patients (36%), thrombosis was related to intravenous catheters or malignant disease. Sensitivity and specificity of duplex ultrasonography were 82% (95% CI, 70% to 93%) and 82% (CI, 72% to 92%), respectively. Venous incompressibility correlated well with thrombosis, whereas only 50% of isolated flow abnormalities proved to be thrombosis-related. **CONCLUSIONS:** Duplex ultrasonography may be the method of choice for initial diagnosis of patients with suspected thrombosis of the upper extremities. However, in patients with isolated flow abnormalities, contrast venography should be performed.

Question: Diagnosis

Study Type: Prospective Cohort

Learner Level: Intermediate

Notes: Pros: Good for discussion of standard Dx issues. Diagnostic gold standard is invasive and imperfect. Can calculate LR. Figure 1 allows for discussion of CONSORT statement Cons: Error in figure 1 numbers

Bhatt, M., et al. (2009). "Prospective validation of the pediatric appendicitis score in a Canadian pediatric emergency department." Acad Emerg Med 16(7): 591-596.

OBJECTIVES: Clinical scoring systems attempt to improve the diagnostic accuracy of pediatric appendicitis. The Pediatric Appendicitis Score (PAS) was the first score created specifically for children and showed excellent performance in the derivation study when administered by pediatric surgeons. The objective was to validate the score in a nonreferred population by emergency physicians (EPs). **METHODS:** A convenience sample of children, 4-18 years old presenting to a pediatric emergency department (ED) with abdominal pain of less than 3 days' duration and in whom the treating physician suspected appendicitis, was prospectively evaluated. Children who were nonverbal, had a previous appendectomy, or had chronic abdominal pathology were excluded. Score components (right lower quadrant and hop tenderness, anorexia, pyrexia, emesis, pain migration, leukocytosis, and neutrophilia) were collected on standardized forms by EPs who were blinded to the scoring system. Interobserver assessments were completed when possible. Appendicitis was defined as appendectomy with positive histology. Outcomes were ascertained by review of the pathology reports from the surgery specimens for children undergoing surgery and by telephone follow-up for children who were discharged home. Sensitivity, specificity, negative predictive value (NPV), and positive predictive value (PPV) were calculated. The overall performance of the score was assessed by a receiver operator characteristic (ROC) curve. **RESULTS:** Of the enrolled children who met inclusion criteria (n = 246), 83 (34%) had pathology-proven appendicitis. Using the single cut-point suggested in the derivation study (PAS 5) resulted in an unacceptably high number of false positives (37.6%). The score's performance improved when two cut-points were used. When children with a PAS of ≤ 4 were discharged home without further investigations, the sensitivity was 97.6% with a NPV of 97.7%. When a PAS of ≥ 8 determined the need for appendectomy, the score's specificity was 95.1% with a PPV of 85.2%. Using this strategy, the negative appendectomy rate would have been 8.8%, the missed appendicitis rate would have been 2.4%, and 41% of imaging investigations would have been avoided. **CONCLUSIONS:** The PAS is a useful tool in the evaluation of children with possible appendicitis. Scores of ≤ 4 help rule out appendicitis, while scores of ≥ 8 help predict appendicitis. Patients with a PAS of 5-7 may need further radiologic evaluation.

Question: Diagnosis

Study Type: Clinical Prediction Rule

Learner Level: Intermediate

Notes: This is prospective study of children with abdominal pain presenting to a Pediatric Emergency Department suspected of having appendicitis. This study was performed to validate a clinical prediction rule to aid in the diagnosis or exclusion of appendicitis. This study provides a nice example of a ROC curve with an accompanying table which clearly demonstrates the interplay between cut-off points, sensitivity and specificity. The table also demonstrates the clinical implications of different cut-off levels (in terms of negative appendectomies and missed diagnoses of appendicitis). This study uses a convenience sample which is good grounds for discussion (e.g. difference in patients enrolled and those missed, real world limitations of performing clinical studies). Additionally, a discussion of the clinical usefulness of this study is worthwhile.

Brownell, R., et al. (2017). "The use of pretest probability increases the value of high-resolution CT in diagnosing usual interstitial pneumonia." Thorax 72(5): 424-429.

BACKGROUND: Recent studies have suggested that non-definitive patterns on high-resolution CT (HRCT) scan provide sufficient diagnostic specificity to forgo surgical lung biopsy in the diagnosis of idiopathic pulmonary fibrosis (IPF). The objective of this study was to determine test characteristics of non-definitive HRCT patterns for identifying histopathological usual interstitial pneumonia (UIP). **METHODS:** Patients with biopsy-proven interstitial lung disease (ILD) and non-definitive HRCT scans were identified from two academic ILD centres. Test characteristics for HRCT patterns as predictors of UIP on surgical lung biopsy were derived and validated in independent cohorts. **RESULTS:** In the derivation cohort, 64/385 (17%) had possible UIP pattern on HRCT; 321/385 (83%) had inconsistent with UIP pattern. 113/385 (29%) patients had histopathological UIP pattern in the derivation cohort. Possible UIP pattern had a specificity of 91.2% (95% CI 87.2% to 94.3%) and a positive predictive value (PPV) of 62.5% (95% CI 49.5% to 74.3%) for UIP pattern on surgical lung biopsy. The addition of age, sex and total traction bronchiectasis score improved the PPV. Inconsistent with UIP pattern demonstrated poor PPV (22.7%, 95% CI 18.3% to 27.7%). HRCT pattern specificity was nearly identical in the validation cohort (92.7%, 95% CI 82.4% to 98.0%). The substantially higher prevalence of UIP pattern in the validation cohort improved the PPV of HRCT patterns. **CONCLUSIONS:** A possible UIP pattern on HRCT has high specificity for UIP on surgical lung biopsy, but PPV is highly dependent on underlying prevalence. Adding clinical and radiographic features to possible UIP pattern on HRCT may provide sufficient probability of histopathological UIP across prevalence ranges to change clinical decision-making.

Question: Diagnosis

Study Type: Clinical Prediction Rule

Learner Level: Advanced

Notes: Derivation and validation of Clinical Prediction Rules for diagnosis.

Budoff, M. J., et al. (2008). "Diagnostic performance of 64-multidetector row coronary computed tomographic angiography for evaluation of coronary artery stenosis in individuals without known coronary artery disease: results from the prospective multicenter ACCURACY (Assessment by Coronary Computed Tomographic Angiography of Individuals Undergoing Invasive Coronary Angiography) trial." *J Am Coll Cardiol* 52(21): 1724-1732.

OBJECTIVES: The purpose of this study was to evaluate the diagnostic accuracy of electrocardiographically gated 64-multidetector row coronary computed tomographic angiography (CCTA) in individuals without known coronary artery disease (CAD). **BACKGROUND:** CCTA is a promising method for detection and exclusion of obstructive coronary artery stenosis. To date, no prospective multicenter trial has evaluated the diagnostic accuracy of 64-multidetector row CCTA in populations with intermediate prevalence of CAD. **METHODS:** We prospectively evaluated subjects with chest pain at 16 sites who were clinically referred for invasive coronary angiography (ICA). CCTAs were scored by consensus of 3 independent blinded readers. The ICAs were evaluated for coronary stenosis based on quantitative coronary angiography (QCA). No subjects were excluded for baseline coronary artery calcium score or body mass index. **RESULTS:** A total of 230 subjects underwent both CCTA and ICA (59.1% male; mean age: 57 +/- 10 years). On a patient-based model, the sensitivity, specificity, and positive and negative predictive values to detect > or =50% or > or =70% stenosis were 95%, 83%, 64%, and 99%, respectively, and 94%, 83%, 48%, 99%, respectively. No differences in sensitivity and specificity were noted for nonobese compared with obese subjects or for heart rates < or =65 beats/min compared with >65 beats/min, whereas calcium scores >400 reduced specificity significantly. **CONCLUSIONS:** In this prospective multicenter trial of chest pain patients without known CAD, 64-multidetector row CCTA possesses high diagnostic accuracy for detection of obstructive coronary stenosis at both thresholds of 50% and 70% stenosis. Importantly, the 99% negative predictive value at the patient and vessel level establishes CCTA as an effective noninvasive alternative to ICA to rule out obstructive coronary artery stenosis. (A Study of Computed Tomography [CT] for Evaluation of Coronary Artery Blockages in Typical or Atypical Chest Pain; NCT00348569).

Question: Diagnosis

Study Type: Diagnostic Tests

Learner Level: Intermediate

Notes: use to assess the validity of studies reporting diagnostic test performance characteristics. Also good to use along with a case scenario to demonstrate likelihood ratios and impact of findings on any one given patient. Intermediate

Bush, K., et al. (1998). "The AUDIT alcohol consumption questions (AUDIT-C): an effective brief screening test for problem drinking. Ambulatory Care Quality Improvement Project (ACQUIP). Alcohol Use Disorders Identification Test." *Arch Intern Med* 158(16): 1789-1795.

OBJECTIVE: To evaluate the 3 alcohol consumption questions from the Alcohol Use Disorders Identification Test (AUDIT-C) as a brief screening test for heavy drinking and/or active alcohol abuse or dependence. **METHODS:** Patients from 3 Veterans Affairs general medical clinics were mailed questionnaires. A random, weighted sample of Health History Questionnaire respondents, who had 5 or more drinks over the past year, were eligible for telephone interviews (N = 447). Heavy drinkers were oversampled 2:1. Patients were excluded if they could not be contacted by telephone, were too ill for interviews, or were female (n = 54). Areas under receiver operating characteristic curves (AUROCs) were used to compare mailed alcohol screening questionnaires (AUDIT-C and full AUDIT) with 3 comparison standards based on telephone interviews: (1) past year heavy drinking (>14 drinks/week or > or =5 drinks/occasion); (2) active alcohol abuse or dependence according to the Diagnostic and Statistical Manual of Mental Disorders, Revised Third Edition, criteria; and (3) either. **RESULTS:** Of 393 eligible patients, 243 (62%) completed AUDIT-C and interviews. For detecting heavy drinking, AUDIT-C had a higher AUROC than the full AUDIT (0.891 vs 0.881; P = .03). Although the full AUDIT performed better than AUDIT-C for detecting active alcohol abuse or dependence (0.811 vs 0.786; P<.001), the 2 questionnaires performed similarly for detecting heavy drinking and/or active abuse or dependence (0.880 vs 0.881). **CONCLUSIONS:** Three questions about alcohol consumption (AUDIT-C) appear to be a practical, valid primary care screening test for heavy drinking and/or active alcohol abuse or dependence.

Question: Diagnosis

Study Type: .

Learner Level: Intermediate

Notes: illustrates LR for various score cutoff thresholds, SpPIn and SnNOuT, and ROC curves; follow-up review by Reinert in 2007 summaries validation of AUDIT-C in other populations

Call, S. A., et al. (2005). "Does this patient have influenza?" *JAMA* 293(8): 987-997.

CONTEXT: Influenza vaccination lowers, but does not eliminate, the risk of influenza. Making a reliable, rapid clinical diagnosis is essential to appropriate patient management that may be especially important during shortages of antiviral agents caused by high demand. **OBJECTIVES:** To systematically review the precision and accuracy of symptoms and signs of influenza. A secondary objective was to review the operating characteristics of rapid diagnostic tests for influenza (results available in <30 min). **DATA SOURCES:** Structured search strategy using MEDLINE (January 1966-September 2004) and subsequent searches of bibliographies of retrieved articles to identify articles describing primary studies dealing with the diagnosis of influenza based on clinical signs and symptoms. The MEDLINE search used the Medical Subject Headings EXP influenza or EXP influenza A virus or EXP influenza A virus human or EXP

influenza B virus and the Medical Subject Headings or terms EXP sensitivity and specificity or EXP medical history taking or EXP physical examination or EXP reproducibility of results or EXP observer variation or symptoms.mp or clinical signs.mp or sensitivity.mp or specificity.mp. STUDY SELECTION: Of 915 identified articles on clinical assessment of influenza-related illness, 17 contained data on the operating characteristics of symptoms and signs using an independent criterion standard. Of these, 11 were eliminated based on 4 inclusion criteria and availability of nonduplicative primary data. DATA EXTRACTION: Two authors independently reviewed and abstracted data for estimating the likelihood ratios (LRs) of clinical diagnostic findings. Differences were resolved by discussion and consensus. DATA SYNTHESIS: No symptom or sign had a summary LR greater than 2 in studies that enrolled patients without regard to age. For decreasing the likelihood of influenza, the absence of fever (LR, 0.40; 95% confidence interval [CI], 0.25-0.66), cough (LR, 0.42; 95% CI, 0.31-0.57), or nasal congestion (LR, 0.49; 95% CI, 0.42-0.59) were the only findings that had summary LRs less than 0.5. In studies limited to patients aged 60 years or older, the combination of fever, cough, and acute onset (LR, 5.4; 95% CI, 3.8-7.7), fever and cough (LR, 5.0; 95% CI, 3.5-6.9), fever alone (LR, 3.8; 95% CI, 2.8-5.0), malaise (LR, 2.6; 95% CI, 2.2-3.1), and chills (LR, 2.6; 95% CI, 2.0-3.2) increased the likelihood of influenza to the greatest degree. The presence of sneezing among older patients made influenza less likely (LR, 0.47; 95% CI, 0.24-0.92). CONCLUSIONS: Clinical findings identify patients with influenza-like illness but are not particularly useful for confirming or excluding the diagnosis of influenza. Clinicians should use timely epidemiologic data to ascertain if influenza is circulating in their communities, then either treat patients with influenza-like illness empirically or obtain a rapid influenza test to assist with management decisions.

Question: Diagnosis

Study Type: Meta-analysis

Learner Level: Intermediate / Advanced

Notes: This meta-analysis is part of the Rational Clinical Exam Series, which reports information on the diagnostic test characteristics of history and physical exam items as well as a limited number of associated diagnostic tests. This article on influenza is timely in that the distribution of flu vaccine was altered by the failure of one European manufacturer to provide expected doses of vaccine to the United States. Thus, this paper was published in a setting of heightened public awareness of risk of influenza. Good discussion points: As with all Rational Clinical Exam articles, this is an evidence summary of diagnostic tests- thus one can discuss both systematic review methodology AND diagnosis, specifically Likelihood ratio. Excellent paragraph under statistical methods (Page 990) that defines likelihood ratio as well as diagnostic odds ratio. Stumbling block may be the large number of items listed in Tables 2 and 3 (p 992 and 993), however a clear difference in data can be noted in patients older than 60 years compared to all comers.

Chung, T., et al. (2006). "Influence of history of heart failure on diagnostic performance and utility of B-type natriuretic peptide testing for acute dyspnea in the emergency department." *Am Heart J* 152(5): 949-955.

BACKGROUND: The aim of this study was to assess the impact of a history of heart failure (HF) on emergency department (ED) B-type natriuretic peptide (BNP) testing and impact of feedback of BNP level to ED physicians. **METHODS:** Admission BNP was measured in 143 patients (mean age 79 +/- 10 years) presenting to the ED with dyspnea. Emergency department physicians scored probability of HF as cause of dyspnea and categorized cause of dyspnea. An independent cardiologist determined cause of dyspnea after chart review. In 83 patients, ED physicians rescored and reclassified patients after BNP measurement and evaluated test utility. **RESULTS:** The area under the receiver operating characteristic curve for BNP diagnosis of HF cause of dyspnea was significantly worse in patients with history of HF than those without (0.74 vs 0.94, $P < .01$) and in those with left ventricular ejection fraction $<50\%$ (0.64 vs 0.87, $P < .05$). A BNP cut point of 100 pg/mL had 100% sensitivity but only 41% specificity for diagnosing acute HF, whereas a cut point of 400 pg/mL had 87% sensitivity and 76% specificity. Emergency department physicians rated BNP useful in 64% of patients, and diagnostic uncertainty was reduced from 53% to 25% ($P < .001$). **CONCLUSION:** B-type natriuretic peptide test performance for diagnosis of dyspnea cause is significantly reduced in patients with a history of HF and must be taken into consideration in the evaluation of such patients in the ED.

Question: Diagnosis

Study Type: .

Learner Level: .

Notes: Among the many BNP articles out there, I like this one because it provides the raw data to allow learners to calculate a likelihood ratio directly from a 2x2 table as opposed to having to interpret an ROC curve. It permits easy calculation of different likelihood ratios for different cutoff points for BNP.

Colli, A., et al. (2006). "Accuracy of ultrasonography, spiral CT, magnetic resonance, and alpha-fetoprotein in diagnosing hepatocellular carcinoma: a systematic review." *Am J Gastroenterol* 101(3): 513-523.

BACKGROUND AND AIM: In patients with chronic liver disease, the accuracy of ultrasound scan (US), spiral computed tomography (CT), magnetic resonance imaging (MRI), and alpha-fetoprotein (AFP) in diagnosing hepatocellular carcinoma (HCC) has never been systematically assessed, and present systematic review was aimed at this issue. **METHODS:** Pertinent cross-sectional studies having as a reference standard pathological examinations of the explanted liver or resected segment(s), biopsies of focal lesion(s), and/or a period of follow-up, were identified using MEDLINE, EMBASE, Cochrane Library, and CancerLit. Pooled sensitivity, specificity, and likelihood ratios (LR) were calculated using the random effect model. Summary receiver operating characteristic (SROC) curve and predefined subgroup analyses were made when indicated. **RESULTS:** The pooled estimates of the 14 US studies were 60% (95% CI 44-76) for sensitivity, 97% (95% CI 95-98) for specificity, 18 (95% CI 8-37) for LR+, and 0.5 (95% CI 0.4-0.6) for LR-; for the 10 CT

studies sensitivity was 68% (95% CI 55-80), specificity 93% (95% CI 89-96), LR+ 6 (95% CI 3-12), and LR- 0.4 (95% CI 0.3-0.6); for the nine MRI studies sensitivity was 81% (95% CI 70-91), specificity 85% (95% CI 77-93), LR+ 3.9 (95% CI 2-7), and LR- 0.3 (95% CI 0.2-0.5). The sensitivity and specificity of AFP varied widely, and this could not be entirely attributed to the threshold effect of the different cutoff levels used. CONCLUSIONS: US is highly specific but insufficiently sensitive to detect HCC in many cirrhotics or to support an effective surveillance program. The operative characteristics of CT are comparable, whereas MRI is more sensitive. High-quality prospective studies are needed to define the actual diagnostic role of AFP.

Question: Diagnosis

Study Type: Systematic Review

Learner Level: Advanced

Notes: Diagnostic tests for hepatocellular carcinoma; Lots of data – you can do calculation of LR from this paper but it will require very much set up first

However, it is a good paper for comparing likelihood ratios, discussing positive and negative LRs; Also can discuss the different cut-offs used for continuous data; Has an accompanying ACP-JC would could easily be used to simplify an exercise on this topic

Collins, J. F., et al. (2005). "Accuracy of screening for fecal occult blood on a single stool sample obtained by digital rectal examination: a comparison with recommended sampling practice." *Ann Intern Med* 142(2): 81-85.

BACKGROUND: Many expert panels recommend colorectal cancer screening for average-risk asymptomatic individuals older than 50 years of age. Recent studies have found that 24% to 64% of primary care providers use only the digital fecal occult blood test (FOBT) as their primary screening test. The effectiveness of a single digital FOBT is unknown. **OBJECTIVE:** To compare the sensitivity and specificity of digital FOBT and the recommended 6-sample at-home FOBT for advanced neoplasia in asymptomatic persons. **DESIGN:** Prospective cohort study. **SETTING:** 13 Veterans Affairs medical centers. **PATIENTS:** 3121 asymptomatic patients 50 to 75 years of age. **INTERVENTION:** 2665 patients had 6-sample at-home FOBT and digital FOBT, followed by complete colonoscopy. **MEASUREMENTS:** We measured the sensitivity of digital and 6-sample FOBT for advanced neoplasia and the specificity for no neoplasia. We calculated predictive values and likelihood ratios for advanced neoplasia, defined as tubular adenomas 10 mm or greater, adenomas with villous histology or high-grade dysplasia, or invasive cancer. **RESULTS:** Of all participants, 96.8% were men; their average age was 63.1 years. The 6-sample FOBT and the single digital FOBT had specificities of 93.9% and 97.5%, respectively, as defined by studying 1656 patients with no neoplasia. Sensitivities for detection of advanced neoplasia in 284 patients were 23.9% for the 6-sample FOBT and 4.9% for the digital FOBT. The likelihood ratio for advanced neoplasia was 1.68 (95% CI, 0.96 to 2.94) for positive results on digital FOBT and 0.98 (CI, 0.95 to 1.01) for negative results. **LIMITATIONS:** Most patients were men. **CONCLUSIONS:** Single digital FOBT is a poor screening method for colorectal neoplasia and cannot be recommended as the only test. When digital FOBT is performed as part of a primary care physical examination, negative results do not decrease the odds of advanced neoplasia. Persons with these results should be offered at-home 6-sample FOBT or another type of screening test.

Question: Diagnosis

Study Type: Prospective Cohort

Learner Level: Beginner / Intermediate

Notes: Prospective cohort in which digital rectal exam is compared to the gold standard of colonoscopy for screening of colon cancer. Important, prevalent issue that comes up (especially in the context of training settings where frequently interns and medical students are "routinely" requested to perform Digital Rectal Exams. Good discussion points: Great paper for discussion of Likelihood Ratios as you can compare the LRs for the 6-sample FOBT and the digital rectal. LR Calculation: Table 3 on page 84 gives you the information to create 2x2 tables. However, the numbers are embedded in the tables in a way you will have to reorganize them into a classic 2x2 table framework to have learners directly calculate LRs. Although this is not hard to do, it is a frequent stumbling block for learners in doing the correct calculation. Applicability questions: great paper for discussion of action thresholds. Regardless of the LR for the different tests, whether or not the use of 6-sample FOBT will be useful in your setting depends on the protocol for screening for colon cancer. Will the 6-sample FOBT be useful for systems that screen with home FOBT cards and subsequent colonoscopy for the positive tests? Will the 6-sample FOBT be useful for systems that screen directly with colonoscopy?

Cotton, P. B., et al. (2004). "Computed tomographic colonography (virtual colonoscopy): a multicenter comparison with standard colonoscopy for detection of colorectal neoplasia." *JAMA* 291(14): 1713-1719.

CONTEXT: Conventional colonoscopy is the best available method for detection of colorectal cancer; however, it is invasive and not without risk. Computed tomographic colonography (CTC), also known as virtual colonoscopy, has been reported to be reasonably accurate in the diagnosis of colorectal neoplasia in studies performed at expert centers. **OBJECTIVE:** To assess the accuracy of CTC in a large number of participants across multiple centers. **DESIGN, SETTING, AND PARTICIPANTS:** A nonrandomized, evaluator-blinded, noninferiority study design of 615 participants aged 50 years or older who were referred for routine, clinically indicated colonoscopy in 9 major hospital centers between April 17, 2000, and October 3, 2001. The CTC was performed by using multislice scanners immediately before standard colonoscopy; findings at colonoscopy were reported before and after segmental unblinding to the CTC results. **MAIN OUTCOME MEASURES:** The sensitivity and specificity of CTC and conventional colonoscopy in detecting participants with lesions sized at least 6 mm. Secondary outcomes included detection of all lesions, detection of advanced lesions, possible technical confounders, participant preferences, and evidence for increasing accuracy with experience. **RESULTS:** A total of 827 lesions were detected in 308 of 600 participants who underwent both procedures; 104 participants had lesions sized at least 6 mm. The sensitivity of CTC for detecting participants with 1 or more lesions sized at least 6 mm was 39.0% (95% confidence interval [CI],

29.6%-48.4%) and for lesions sized at least 10 mm, it was 55.0% (95% CI, 39.9%-70.0%). These results were significantly lower than those for conventional colonoscopy, with sensitivities of 99.0% (95% CI, 97.1%-99.9%) and 100%, respectively. A total of 496 participants were without any lesion sized at least 6 mm. The specificity of CTC and conventional colonoscopy for detecting participants without any lesion sized at least 6 mm was 90.5% (95% CI, 87.9%-93.1%) and 100%, respectively, and without lesions sized at least 10 mm, 96.0% (95% CI, 94.3%-97.6%) and 100%, respectively. Computed tomographic colonography missed 2 of 8 cancers. The accuracy of CTC varied considerably between centers and did not improve as the study progressed. Participants expressed no clear preference for either technique. CONCLUSIONS: Computed tomographic colonography by these methods is not yet ready for widespread clinical application. Techniques and training need to be improved.

Question: Screening

Study Type: Non-randomized, evaluator blinded, non-inferiority study

Learner Level: Intermediate

Notes: Outstanding methodology; Important clinical question – readiness of virtual colonoscopy for prime-time; Good discussion points: Excellent data for calculation of likelihood ratios at different cut-offs of size of polyps, answers given in ACP journal club EBM review of paper; Good review for sensitivity and specificity; Good discussion re statistical sensitivity and specificity vs clinical utility.

De Smet, D., et al. (2010). "Use of likelihood ratios improves interpretation of laboratory testing for pulmonary sarcoidosis." American Journal of Clinical Pathology 134(6): 939-947.

Laboratory tests for pulmonary sarcoidosis (percentage lymphocytes and CD4/CD8 ratio in bronchoalveolar lavage fluid and serum angiotensin-converting enzyme activity) lack sensitivity and specificity. In a retrospective study of 153 subjects under suspicion of pulmonary sarcoidosis (36 cases and 117 patients with other diseases [control patients]), we defined likelihood ratios (LRs) for rationally selected result intervals of these tests, which improve clinical interpretation as compared with dichotomous interpretation based on a single cutoff value. By using logistic regression analysis, we further integrated the 3 individual tests into a unified algorithm that could rule out diagnosis in 57 (48.7%) of the 177 control subjects and confirm diagnosis in 12 (33%) of the 36 pulmonary sarcoidosis cases. We conclude that use of LRs improves interpretation of laboratory tests for pulmonary sarcoidosis. In addition, we present a prediction algorithm based on the combination of laboratory tests that helps clinicians confirm or exclude diagnosis in almost half of the study population.

Question: Diagnosis

Study Type: Cohort Study

Learner Level: Beginner

Notes: Multiple cut-offs for ACE level are good for teaching how to calculate likelihood ratio.

Demirer, E., et al. (2012). "Predictive models for tuberculous pleural effusions in a high tuberculosis prevalence region." Lung 190(2): 239-248.

BACKGROUND: Patients with pleural effusions who reside in geographic areas with a high prevalence of tuberculosis frequently have similar clinical manifestations of other diseases. The aim of our study was to develop a simple but accurate clinical score for differential diagnosis of tuberculous pleural effusion (TPE) from non-TB pleural effusion (NTPE). METHODS: This was an unblinded, prospective study of Turkish patients 18 years of age or older with pleural effusion of indeterminate etiology conducted from June 2003 to June 2005. Unconditional logistic regression models were used to discriminate TPE cases from NTPE cases. Standard errors for the area under the curve (AUC) were calculated using the Mann-Whitney method. Data were statistically significant if two-tailed $P < 0.05$. RESULTS: A total of 63.3% (157/248) of the patients had TPE while 36.7% (91/248) of the patients had other etiologies for pleural effusions. We were able to provide a predictive model of TPE that included age < 47 years and either pleural fluid adenosine deaminase enzyme (PADA) > 35 U/l or pleural serum protein ratio > 0.710 . However, only the combination of age < 47 and PADA > 35 U/l was significant (odds ratio [OR]: 7.46; 95% confidence interval [CI]: 3.99-13.96). The generated summary score (range = 0-6) was significantly predictive of TPE (OR: 2.91; 95% CI: 2.18-3.89) and with high AUC (0.79). CONCLUSION: We propose an affordable model that includes age < 47 years and PADA > 35 U/l for timely diagnosis of TPE in geographical regions with a high prevalence of TB.

Question: Diagnosis

Study Type: Prospective

Learner Level: .

Notes: Diagnosis

de-Wahl Granelli, A., et al. (2009). "Impact of pulse oximetry screening on the detection of duct dependent congenital heart disease: a Swedish prospective screening study in 39,821 newborns." BMJ 338: a3037.

OBJECTIVE: To evaluate the use of pulse oximetry to screen for early detection of life threatening congenital heart disease. DESIGN: Prospective screening study with a new generation pulse oximeter before discharge from well baby nurseries in West Gotland.

Cohort study comparing the detection rate of duct dependent circulation in West Gotaland with that in other regions not using pulse oximetry screening. Deaths at home with undetected duct dependent circulation were included. SETTING: All 5 maternity units in West Gotaland and the supraregional referral centre for neonatal cardiac surgery. PARTICIPANTS: 39,821 screened babies born between 1 July 2004 and 31 March 2007. Total duct dependent circulation cohorts: West Gotaland n=60, other referring regions n=100. MAIN OUTCOME MEASURES: Sensitivity, specificity, positive and negative predictive values, and likelihood ratio for pulse oximetry screening and for neonatal physical examination alone. RESULTS: In West Gotaland 29 babies in well baby nurseries had duct dependent circulation undetected before neonatal discharge examination. In 13 cases, pulse oximetry showed oxygen saturations <or=90%, and (in accordance with protocol) clinical staff were immediately told of the results. Of the remaining 16 cases, physical examination alone detected 10 (63%). Combining physical examination with pulse oximetry screening had a sensitivity of 24/29 (82.8% (95% CI 64.2% to 95.2%)) and detected 100% of the babies with duct dependent lung circulation. Five cases were missed (all with aortic arch obstruction). False positive rate with pulse oximetry was substantially lower than that with physical examination alone (69/39 821 (0.17%) v 729/38 413 (1.90%), P<0.0001), and 31/69 of the "false positive" cases with pulse oximetry had other pathology. Thus, referral of all cases with positive oximetry results for echocardiography resulted in only 2.3 echocardiograms with normal cardiac findings for every true positive case of duct dependent circulation. In the cohort study, the risk of leaving hospital with undiagnosed duct dependent circulation was 28/100 (28%) in other referring regions versus 5/60 (8%) in West Gotaland (P=0.0025, relative risk 3.36 (95% CI 1.37 to 8.24)). In the other referring regions 11/25 (44%) of babies with transposition of the great arteries left hospital undiagnosed versus 0/18 in West Gotaland (P=0.0010), and severe acidosis at diagnosis was more common (33/100 (33%) v 7/60 (12%), P=0.0025, relative risk 2.8 (1.3 to 6.0)). Excluding premature babies and Norwood surgery, babies discharged without diagnosis had higher mortality than those diagnosed in hospital (4/27 (18%) v 1/110 (0.9%), P=0.0054). No baby died from undiagnosed duct dependent circulation in West Gotaland versus five babies from the other referring regions. CONCLUSION: Introducing pulse oximetry screening before discharge improved total detection rate of duct dependent circulation to 92%. Such screening seems cost neutral in the short term, but the probable prevention of neurological morbidity and reduced need for preoperative neonatal intensive care suggest that such screening will be cost effective long term.

Question: Diagnosis

Study Type: Prospective Cohort Study

Learner Level: Intermediate

Notes: Can be used to demonstrate the very high likelihood ratios necessary when screening for relatively rare conditions, and the approach of using the databases available within a coordinated healthcare system to do follow-up on negative screens in situations where the referent standard (echocardiography) cannot be feasibly be applied to all. Excellent consecutive enrollment, blinding complicated (could not blind interpreters dealing with babies with particularly low O2 sats, 90%), although unlikely to affect outcome.

Flaherty, K. R., et al. (2004). "Idiopathic interstitial pneumonia: what is the effect of a multidisciplinary approach to diagnosis?" Am J Respir Crit Care Med 170(8): 904-910.

Current guidelines recommend that the clinician, radiologist, and pathologist work together to establish a diagnosis of idiopathic interstitial pneumonia. Three clinicians, two radiologists, and two pathologists reviewed 58 consecutive cases of suspected idiopathic interstitial pneumonia. Each participant was provided information in a sequential manner and was asked to record their diagnostic impression and level of confidence at each step. Interobserver agreement improved from the beginning to the end of the review. After the presentation of histopathologic information, radiologists changed their diagnostic impression more often than did clinicians. In general, as more information was provided the confidence level for a given diagnosis improved, and the diagnoses rendered with a high level of confidence were more likely congruent with the final pathologic consensus diagnosis. The final consensus pathologist diagnosis was idiopathic pulmonary fibrosis in 30 cases. Clinicians identified 75% and radiologists identified 48% of these cases before presentation of the histopathologic information. Histopathologic information has the greatest impact on the final diagnosis, especially when the initial clinical/radiographic diagnosis is not idiopathic pulmonary fibrosis. We conclude that dynamic interactions between clinicians, radiologists, and pathologists improve interobserver agreement and diagnostic confidence.

Question: Diagnosis

Study Type: .

Learner Level: .

Notes: Diagnosis

Froehlich, T. E., et al. (1998). "Screening for dementia in the outpatient setting: the time and change test." J Am Geriatr Soc 46(12): 1506-1511.

OBJECTIVE: To develop and validate the Time and Change (T&C) test, a simple, standardized method for detecting dementia in a diverse older outpatient population with varying levels of education. **DESIGN:** A prospective cohort validation study. **SETTING:** Two outpatient clinics at an urban teaching hospital. **PARTICIPANTS:** The concurrent validation sample consisted of 100 consecutive outpatients 70 years of age or older who were 58% non-white and had a 16% dementia prevalence rate and educational levels ranging from 0 to 17+ years. Reliability was tested in a sample of 42 consecutive outpatients 75 years of age or older with a 36% dementia prevalence rate. **MEASUREMENTS:** T&C ratings were validated against a reference standard based on the Blessed Dementia Rating Scale and the Mini-Mental State Examination. Reliability, contribution to physician recognition of dementia, and ease of use were assessed. **RESULTS:** In the outpatient setting, the T&C had a sensitivity of 63%, specificity of 96%, a negative

predictive value of 93%, a positive predictive value of 77%, and test-retest and inter-observer reliability agreement rates of 95% and 100%, respectively. When T&C results were added to the physician's documentation of dementia, the number of missed cases decreased from 44% to 19%, and the number of overcalled cases decreased by 100%. When timed cut points were added, the T&C test had a sensitivity of 94 to 100%, specificity of 37 to 46%, negative predictive value of 98 to 100%, positive predictive value of 23 to 25%, and test-retest and inter-observer agreement rates of 82% and 70 to 75%, respectively. CONCLUSION: The T&C test is a simple, accurate, reliable, performance-based tool that can improve physician ability to recognize dementia in diverse outpatient populations.

Question: Diagnosis

Study Type: Cross-Sectional

Learner Level: Beginner

Notes: This is another simple diagnosis article that is great for calculating likelihood ratios. Also good in that it's an outpatient intervention that can be done fairly easily--nice example of how EBM can be used in "real-world" settings.

Garcia-Zamalloa, A. and J. Taboada-Gomez (2012). "Diagnostic accuracy of adenosine deaminase and lymphocyte proportion in pleural fluid for tuberculous pleurisy in different prevalence scenarios." [PloS One](#) 7(6): e38729.

BACKGROUND: Tuberculous pleural effusion (TPE) is a paucibacillary manifestation of tuberculosis, so isolation of Mycobacterium tuberculosis is difficult, biomarkers being an alternative for diagnosis. Adenosine deaminase (ADA) is the most cost-effective pleural fluid marker and is routinely used in high prevalence settings, whereas its value is questioned in areas with low prevalence. The lymphocyte proportion (LP) is known to increase the specificity of ADA for this diagnosis. We analyse the diagnostic usefulness of ADA alone and the combination of ADA \geq 40 U/l (ADA(40)) and LP \geq 50% (LP(50)) in three different prevalence scenarios over 11 years in our area. MATERIALS AND METHODS: Biochemistry, cytology and microbiology studies from 472 consecutive pleural fluid samples were retrospectively analyzed. ADA and differential cell count were determined in all samples. We established three different prevalence periods, based on percentage of pleural effusion cases diagnosed as tuberculosis: 1998-2000 (31.3%), 2001-2004 (11.8%), and 2005-2008 (7.4%). ROC curves, dispersion diagrams and pre/post-test probability graphs were produced. TPE accounted for 73 episodes (mean prevalence: 15.5%). The sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV) for ADA(40) were 89%, 92.7%, 69.2% and 97.9%, respectively. For ADA(40)+LP(50) the specificity and PPV increased (98.3% and 90%) with hardly any decrease in the sensitivity or NPV (86.3% and 97.5%). No relevant differences were observed between the three study periods. CONCLUSIONS/SIGNIFICANCE: ADA remains useful for the diagnosis of TPE even in low-to-intermediate prevalence scenarios when combined with the lymphocyte proportion.

Question: Diagnosis

Study Type: .

Learner Level: .

Notes: Diagnosis

Grayson, M. L., et al. (1995). "Probing to bone in infected pedal ulcers. A clinical sign of underlying osteomyelitis in diabetic patients." [JAMA](#) 273(9): 721-723.

OBJECTIVE--To assess a bedside technique for diagnosing osteomyelitis. DESIGN--We prospectively assessed infected pedal ulcers for detectable bone by probing with a sterile, blunt, stainless steel probe. We then examined the relationship between detection of bone and the presence or absence of osteomyelitis that was defined histopathologically and/or clinically. SETTING--A tertiary care center. PATIENTS--Seventy-five hospitalized diabetic patients with a total of 76 infected foot ulcers were studied. RESULTS--Osteomyelitis was diagnosed in 50 instances (66%) and was excluded in 26 instances. Bone was detected by probing in 33 of 50 ulcers with contiguous osteomyelitis; in contrast, bone was probed in only four of 26 ulcers without contiguous osteomyelitis ($P < .001$). Bone detected on probing was visible in only three instances. Palpating bone on probing the pedal ulcer had a sensitivity of 66% for osteomyelitis, a specificity of 85%, a positive predictive value of 89%, and a negative predictive value of 56%. CONCLUSIONS--Palpation of bone in the depths of infected pedal ulcers in patients with diabetes is strongly correlated with the presence of underlying osteomyelitis. If bone is palpated on probing, specialized roentgenographic and radionuclide tests to diagnose osteomyelitis are unnecessary. Probing for bone should be included in the initial assessment of all diabetic patients with infected pedal ulcers.

Question: Diagnosis

Study Type: Prospective Cohort

Learner Level: Beginner / Intermediate

Notes: Strong methodology paper on physical examination as a diagnostic test; Good discussion points; Good for discussion of 'kappa' (inter-observer agreement); One glitch is that you have to pull the raw data out of the text in order to calculate Likelihood ratios or use the sensitivity and specificity in table on p 722. However, it can be done... (one might consider looking for the ACP journal club summary on this paper for guidance)

Guyatt, G. H., et al. (1990). "Diagnosis of iron-deficiency anemia in the elderly." Am J Med 88(3): 205-209.

PURPOSE: To determine the value of serum ferritin, mean cell volume, transferrin saturation, and free erythrocyte protoporphyrin in the diagnosis of iron-deficiency anemia in the elderly. **PATIENTS AND METHODS:** We prospectively studied consecutive eligible and consenting anemic patients over the age of 65 years, who underwent blood tests and bone marrow aspiration. The study consisted of 259 inpatients and outpatients at two community hospitals in whom a complete blood count processed by the hospital laboratory demonstrated previously undiagnosed anemia (men: hemoglobin level less than 12 g/dL; women: hemoglobin level less than 11.0 g/dL). **RESULTS:** Thirty-six percent of our patients had no demonstrable marrow iron and were classified as being iron-deficient. The serum ferritin was the best test for distinguishing those with iron deficiency from those who were not iron-deficient. No other test added clinically important information. The likelihood ratios associated with the serum ferritin level were as follows: greater than 100 micrograms/L, 0.13; greater than 45 micrograms/L but less than or equal to 100 micrograms/L, 0.46; greater than 18 micrograms/L but less than or equal to 45 micrograms/L, 3.12; and less than or equal to 18 micrograms/L, 41.47. These results indicate that values up to 45 micrograms/L increase the likelihood of iron deficiency, whereas values over 45 micrograms/L decrease the likelihood of iron deficiency. Seventy-two percent of those who were not iron-deficient had serum ferritin values greater than 100 micrograms/L, and in populations with a prevalence of iron deficiency of less than 40%, values of greater than 100 micrograms/L reduce the probability of iron deficiency to under 10%. Fifty-five percent of the iron-deficient patients had serum ferritin values of less than 18 micrograms/L, and in populations with a prevalence of iron deficiency of greater than 20%, values of less than 18 micrograms/L increase the probability of iron deficiency to over 95%.

Question: Diagnosis

Study Type: Prospective Cohort

Learner Level: Beginner / Intermediate; Classic paper

Notes: Classic Diagnosis paper for teaching likelihood ratios. This paper, published the same year as PLOPED (above) serves as the other 'golden' paper for teaching LRs. See all above notes: strength of methods and great results that lend themselves to perfect discussion of the value of likelihood ratios for ferritin. From a clinical point of view, ferritin remains an important diagnostic test for evaluation of anemia. One of our all time favorite teaching papers

Hart, R. (2006). "A simple risk score predicted 7-day stroke risk after transient ischemic attack." ACP Journal Club 144(1): 24.

Question: Clinical Prediction Guideline

Study Type: Observational

Learner Level: Intermediate

Notes: Highly applicable to internal medicine; No illustration of variability in results of cohorts

Henschke, C. I., et al. (2006). "Survival of patients with stage I lung cancer detected on CT screening." N Engl J Med 355(17): 1763-1771.

BACKGROUND: The outcome among patients with clinical stage I cancer that is detected on annual screening using spiral computed tomography (CT) is unknown. **METHODS:** In a large collaborative study, we screened 31,567 asymptomatic persons at risk for lung cancer using low-dose CT from 1993 through 2005, and from 1994 through 2005, 27,456 repeated screenings were performed 7 to 18 months after the previous screening. We estimated the 10-year lung-cancer-specific survival rate among participants with clinical stage I lung cancer that was detected on CT screening and diagnosed by biopsy, regardless of the type of treatment received, and among those who underwent surgical resection of clinical stage I cancer within 1 month. A pathology panel reviewed the surgical specimens obtained from participants who underwent resection. **RESULTS:** Screening resulted in a diagnosis of lung cancer in 484 participants. Of these participants, 412 (85%) had clinical stage I lung cancer, and the estimated 10-year survival rate was 88% in this subgroup (95% confidence interval [CI], 84 to 91). Among the 302 participants with clinical stage I cancer who underwent surgical resection within 1 month after diagnosis, the survival rate was 92% (95% CI, 88 to 95). The 8 participants with clinical stage I cancer who did not receive treatment died within 5 years after diagnosis. **CONCLUSIONS:** Annual spiral CT screening can detect lung cancer that is curable.

Question: Screening

Study Type: Cohort study

Learner Level: Intermediate

Notes: great to contrast cohort vs RCT (Aberle) data for screening

Hlibczuk, V., et al. (2010). "Diagnostic accuracy of noncontrast computed tomography for appendicitis in adults: a systematic review." Ann Emerg Med 55(1): 51-59 e51.

STUDY OBJECTIVE: We seek to determine the diagnostic test characteristics of noncontrast computed tomography (CT) for appendicitis in the adult emergency department (ED) population. **METHODS:** We conducted a search of MEDLINE, EMBASE, the Cochrane Library, and the bibliographies of previous systematic reviews. Included studies assessed the diagnostic accuracy of

noncontrast CT for acute appendicitis in adults by using the final diagnosis at surgery or follow-up at a minimum of 2 weeks as the reference standard. Studies were included only if the CT was completed using a multislice helical scanner. Two authors independently conducted the relevance screen of titles and abstracts, selected studies for the final inclusion, extracted data, and assessed study quality. Consensus was reached by conference, and any disagreements were adjudicated by a third reviewer. Unenhanced CT test performance was assessed with summary receiver operating characteristic curve analysis, with independently pooled sensitivity and specificity values across studies. RESULTS: The search yielded 1,258 publications; 7 studies met the inclusion criteria and provided a sample of 1,060 patients. The included studies were of high methodological quality with respect to appropriate patient spectrum and reference standard. Our pooled estimates for sensitivity and specificity were 92.7% (95% confidence interval 89.5% to 95.0%) and 96.1% (95% confidence interval 94.2% to 97.5%), respectively; the positive likelihood ratio=24 and the negative likelihood ratio=0.08. CONCLUSION: We found the diagnostic accuracy of noncontrast CT for the diagnosis of acute appendicitis in the adult population to be adequate for clinical decisionmaking in the ED setting.

Question: Diagnosis

Study Type: Systematic Review

Learner Level: .

Notes: Non-contrast appi. While most people think of SRs as involving therapy issues, there are also diagnostic SRs which offer some unique challenges from a critical appraisal perspective. This article can also serve as great substrate for a GRADE exercise or a Knowledge Translation effort.

Imperiale, T. F., et al. (2014). "Multitarget stool DNA testing for colorectal-cancer screening." *N Engl J Med* 370(14): 1287-1297.

BACKGROUND: An accurate, noninvasive test could improve the effectiveness of colorectal-cancer screening. METHODS: We compared a noninvasive, multitarget stool DNA test with a fecal immunochemical test (FIT) in persons at average risk for colorectal cancer. The DNA test includes quantitative molecular assays for KRAS mutations, aberrant NDRG4 and BMP3 methylation, and beta-actin, plus a hemoglobin immunoassay. Results were generated with the use of a logistic-regression algorithm, with values of 183 or more considered to be positive. FIT values of more than 100 ng of hemoglobin per milliliter of buffer were considered to be positive. Tests were processed independently of colonoscopic findings. RESULTS: Of the 9989 participants who could be evaluated, 65 (0.7%) had colorectal cancer and 757 (7.6%) had advanced precancerous lesions (advanced adenomas or sessile serrated polyps measuring ≥ 1 cm in the greatest dimension) on colonoscopy. The sensitivity for detecting colorectal cancer was 92.3% with DNA testing and 73.8% with FIT ($P=0.002$). The sensitivity for detecting advanced precancerous lesions was 42.4% with DNA testing and 23.8% with FIT ($P<0.001$). The rate of detection of polyps with high-grade dysplasia was 69.2% with DNA testing and 46.2% with FIT ($P=0.004$); the rates of detection of serrated sessile polyps measuring 1 cm or more were 42.4% and 5.1%, respectively ($P<0.001$). Specificities with DNA testing and FIT were 86.6% and 94.9%, respectively, among participants with nonadvanced or negative findings ($P<0.001$) and 89.8% and 96.4%, respectively, among those with negative results on colonoscopy ($P<0.001$). The numbers of persons who would need to be screened to detect one cancer were 154 with colonoscopy, 166 with DNA testing, and 208 with FIT. CONCLUSIONS: In asymptomatic persons at average risk for colorectal cancer, multitarget stool DNA testing detected significantly more cancers than did FIT but had more false positive results. (Funded by Exact Sciences; ClinicalTrials.gov number, NCT01397747).

Question: Diagnosis

Study Type: .

Learner Level: Intermediate/Advanced

Notes: Intermediate to advanced Diagnosis article – good for discussion on expectation, overdiagnosis, lead-time and length-time biases. Also NNS.

Julayanont, P., et al. (2014). "Montreal Cognitive Assessment Memory Index Score (MoCA-MIS) as a predictor of conversion from mild cognitive impairment to Alzheimer's disease." *J Am Geriatr Soc* 62(4): 679-684.

OBJECTIVES: To assess the usefulness of the Montreal Cognitive Assessment (MoCA) total score (MoCA-TS) and Memory Index Score (MoCA-MIS) in predicting conversion to Alzheimer's disease (AD) in individuals with mild cognitive impairment (MCI). DESIGN: Retrospective chart review. SETTING: Community-based memory clinic. PARTICIPANTS: Individuals meeting Petersen's MCI criteria ($N = 165$). MEASUREMENTS: Baseline MoCA scores at MCI diagnosis were collected from charts of eligible individuals with MCI, and MoCA-TS, MoCA-MIS, and a cognitive domain index score were calculated to assess their prognostic value in predicting conversion to AD. RESULTS: One hundred fourteen participants progressed to AD (MCI-AD), and 51 did not (nonconverters; MCI-NC); 90.5% of participants with MCI with a MoCA-TS less than 20/30 and a MoCA-MIS less than 7/15 at baseline converted to AD within the average follow-up period of 18 months, compared with 52.7% of participants with MCI above the cutoffs on both scores. Individuals with multiple-domain amnesic MCI had the highest AD conversion rates (73.9%). CONCLUSION: Identifying individuals with MCI at high risk of conversion to AD is important clinically and for selecting appropriate subjects for therapeutic trials. Individuals with MCI with a low MoCA-TS and a low newly devised memory index score (MoCA-MIS) are at greater risk of short-term conversion to AD.

Question: Diagnosis

Study Type: Retro Chart Review

Learner Level: Beginner

Notes: *Duplicate of 2005 article of same title, not sure if has update so included. Diagnosis – Simple cognitive testing done in the clinic in patients with mild cognitive impairment to predict conversion to Alzheimer's. Easy numbers for calculating multi-level likelihood

Koppel, R. I., et al. (2003). "Effectiveness of pulse oximetry screening for congenital heart disease in asymptomatic newborns." Pediatrics 111(3): 451-455.

OBJECTIVE: To determine the sensitivity, specificity, predictive value, and accuracy of a program of pulse oximetry screening of asymptomatic newborns for critical congenital cardiovascular malformation (CCVM). **METHODS:** Pulse oximetry was performed on asymptomatic newborns in the well-infant nurseries of 2 hospitals. Cardiac ultrasound was performed on infants with positive screens (saturation \leq 95% at $>$ 24 hours). Data regarding true and false positives as well as negatives were collected and analyzed. **RESULTS:** Oximetry was performed on 11 281 asymptomatic newborns, and 3 cases of CCVM were detected (total anomalous pulmonary venous return x2, truncus arteriosus). During the study interval, there were 9 live births of infants with CCVM from a group of 15 fetuses with CCVM detected by fetal echocardiography. Six infants with CCVM were symptomatic before screening. There was 1 false-positive screen. Two infants with negative screens were readmitted (coarctation, hypoplastic left pulmonary artery with aorto-pulmonary collaterals). Other cardiac diagnoses in the database search were nonurgent, including cases of patent foramen ovale, peripheral pulmonic stenosis, and ventricular septal defect. The prevalence of critical CCVM among all live births was 1 in 564 and among the screened population was 1 in 2256 (sensitivity: 60%; specificity: 99.95%; positive predictive value: 75%; negative predictive value: 99.98%; accuracy: 99.97%). **CONCLUSIONS:** This screening test is simple, noninvasive, and inexpensive and can be administered in conjunction with state-mandated screening. The false-negative screen patients had lesions not amenable to detection by oximetry. The sensitivity, specificity, and predictive value in this population are satisfactory, indicating that screening should be applied to larger populations, particularly where lower rates of fetal detection result in increased CCVM prevalence in asymptomatic newborns.

Question: Diagnosis

Study Type: Prospective Cohort

Learner Level: Beginner / Intermediate

Notes: Strong methodology paper; Very good for discussion about likelihood ratios (impressive numbers will make an impact)

Kuppermann, N., et al. (2009). "Identification of children at very low risk of clinically-important brain injuries after head trauma: a prospective cohort study." Lancet 374(9696): 1160-1170.

BACKGROUND: CT imaging of head-injured children has risks of radiation-induced malignancy. Our aim was to identify children at very low risk of clinically-important traumatic brain injuries (ctTBI) for whom CT might be unnecessary. **METHODS:** We enrolled patients younger than 18 years presenting within 24 h of head trauma with Glasgow Coma Scale scores of 14-15 in 25 North American emergency departments. We derived and validated age-specific prediction rules for ctTBI (death from traumatic brain injury, neurosurgery, intubation $>$ 24 h, or hospital admission $>$ or=2 nights). **FINDINGS:** We enrolled and analysed 42 412 children (derivation and validation populations: 8502 and 2216 younger than 2 years, and 25 283 and 6411 aged 2 years and older). We obtained CT scans on 14 969 (35.3%); ctTBIs occurred in 376 (0.9%), and 60 (0.1%) underwent neurosurgery. In the validation population, the prediction rule for children younger than 2 years (normal mental status, no scalp haematoma except frontal, no loss of consciousness or loss of consciousness for less than 5 s, non-severe injury mechanism, no palpable skull fracture, and acting normally according to the parents) had a negative predictive value for ctTBI of 1176/1176 (100.0%, 95% CI 99.7-100.0) and sensitivity of 25/25 (100%, 86.3-100.0). 167 (24.1%) of 694 CT-imaged patients younger than 2 years were in this low-risk group. The prediction rule for children aged 2 years and older (normal mental status, no loss of consciousness, no vomiting, non-severe injury mechanism, no signs of basilar skull fracture, and no severe headache) had a negative predictive value of 3798/3800 (99.95%, 99.81-99.99) and sensitivity of 61/63 (96.8%, 89.0-99.6). 446 (20.1%) of 2223 CT-imaged patients aged 2 years and older were in this low-risk group. Neither rule missed neurosurgery in validation populations. **INTERPRETATION:** These validated prediction rules identified children at very low risk of ctTBIs for whom CT can routinely be obviated. **FUNDING:** The Emergency Medical Services for Children Programme of the Maternal and Child Health Bureau, and the Maternal and Child Health Bureau Research Programme, Health Resources and Services Administration, US Department of Health and Human Services.

Question: Diagnosis

Study Type: Clinical Prediction Rule

Learner Level: Intermediate

Notes: This is a well-designed large, multi-centered, prospective study of a clinical decision rule for the use of Head CTs in children with minor head injury. One could discuss the validity criteria for Diagnostic studies or Clinical Prediction Rules. This study both derives and validates the prediction rule allowing discussion of the importance of validating a prediction in another population of patients. This study also lends itself to discussion of test characteristics (sensitivity, specificity, NPV and PPV) including likelihood ratios.

Lacasse, Y., et al. (2003). "Clinical diagnosis of hypersensitivity pneumonitis." Am J Respir Crit Care Med 168(8): 952-958.

The diagnosis of hypersensitivity pneumonitis (HP) is difficult and often relies on histopathology. Our objective was to identify diagnostic criteria and to develop a clinical prediction rule for this disease. Consecutive patients presenting a condition for which HP was considered in the differential diagnosis underwent a program of simple standardized diagnostic procedures. High-resolution computed tomography scan and bronchoalveolar lavage (BAL) defined the presence or absence of HP. Patients underwent surgical lung biopsy when the computed tomography scan, BAL, and other diagnostic procedures failed to yield a diagnosis. A cohort of 400 patients (116 with HP, 284 control subjects) provided data for the rule derivation. Six significant predictors of HP were identified: (1) exposure to a known offending antigen, (2) positive precipitating antibodies to the offending antigen, (3) recurrent episodes of symptoms, (4) inspiratory crackles on physical examination, (5) symptoms occurring 4 to 8 hours after exposure, (6) and weight loss. The area under the receiver operating characteristic curve was 0.93 (95% confidence interval: 0.90-0.95). The rule retained its accuracy when validated in a separate cohort of 261 patients. The diagnosis of HP can often be made or rejected with confidence, especially in areas of high or low prevalence, respectively, without BAL or biopsy.

Question: Diagnosis

Study Type: Clinical Prediction Rule

Learner Level: Advanced

Notes: Derivation and validation of Clinical Prediction Rules for diagnosis.

Lateef, T., et al. (2013). "Validation of a migraine interview for children and adolescents." *Pediatrics* 131(1): e96-102.

OBJECTIVE: To date there are no structured interviews to ascertain the diagnostic criteria for headache in children. The objective of this study was to assess the validity of the Diagnostic Interview of Headache Syndromes-Child Version (DIHS-C), which was developed at the National Institute of Mental Health for a community-based family study of headache syndromes and comorbid disorders. **METHODS:** The DIHS-C is a fully structured diagnostic interview composed of an open-ended clinical history, modules with key symptoms for each of the major headache subtypes, and associated impairment, duration, frequency, course, and treatment. This article presents the validation of the interview in a sample of 104 children evaluated as part of a community-based family study of migraine. **RESULTS:** The sensitivity of interview diagnosis compared with an expert neurologist's diagnosis of migraine was 98%, and the specificity was 61%. Similar levels of sensitivity and specificity were found by gender and age of the children. **CONCLUSIONS:** The DIHS-C provides a new tool that can enhance the reliability of pediatric diagnoses in both clinical and community settings.

Question: Diagnosis

Study Type: .

Learner Level: Beginner

Notes: Diagnosis. a) it deals with a relatively common disease (migraine), b) it shows that clinicians' questions to patients are diagnostic tests, and c) illustrated the limitations of testing diagnostic tools for conditions in which there is no "hard" gold standard.

Lee, G. M. and M. B. Harper (1998). "Risk of bacteremia for febrile young children in the post-Haemophilus influenzae type b era." *Archives of Pediatrics and Adolescent Medicine* 152(7): 624-628.

OBJECTIVES: To determine the risk for bacteremia, in the post-Haemophilus influenzae type b era, in a prospective cohort of well-appearing febrile children 3 to 36 months of age with no obvious source of infection; and to compare the predictive abilities of objective criteria in identification of children with occult pneumococcal bacteremia from those at risk. **DESIGN:** All children seen from 1993 through 1996, 3 to 36 months of age with a temperature of 39.0 degrees C or higher, no identified source of infection (except otitis media), and discharged to home were considered to be at risk for occult bacteremia and included in the study. **SETTING:** Urban pediatric emergency department. **RESULTS:** Of 199868 patient visits to the emergency department, 1911 children were considered to be at risk for occult bacteremia. Blood cultures were obtained from 9465 (79%). A total of 149 blood cultures contained pathogenic organisms, indicating a rate of occult bacteremia of 1.57% (95% confidence intervals: 1.32%-1.83%). White blood cell count and absolute neutrophil count were the best predictors for occult pneumococcal bacteremia. Using a white blood cell count cutoff value of 15 cells x 10(9)/L (sensitivity, 86%; specificity, 77%; and positive predictive value, 5.1%) would result in the treatment of approximately 19 nonbacteremic children for each bacteremic child treated. **CONCLUSIONS:** The prevalence of occult bacteremia in children 3 to 36 months old with temperatures of 39.0 degrees C or higher and no obvious source of infection is 1.6%. The white blood cell and absolute neutrophil counts are the most accurate predictors of occult pneumococcal bacteremia and when available should be used if presumptive antibiotic therapy is being considered.

Question: Diagnosis

Study Type: Cohort

Learner Level: Intermediate

Notes: Good methodology Allows calculation of multi-level likelihood ratios. Provides some of the background evidence for current practice. Also allows learner to see that tests are not dichotomous. This is a paper that is similarly useful to the classic Guyatt paper on Fe deficiency.

Lieberman, D. A. and D. G. Weiss (2001). "One-time screening for colorectal cancer with combined fecal occult-blood testing and examination of the distal colon." N Engl J Med 345(8): 555-560.

BACKGROUND: Fecal occult-blood testing and sigmoidoscopy have been recommended for screening for colorectal cancer, but the sensitivity of such combined testing for detecting neoplasia is uncertain. At 13 Veterans Affairs medical centers, we performed colonoscopy to determine the prevalence of neoplasia and the sensitivity of one-time screening with a fecal occult-blood test plus sigmoidoscopy. **METHODS:** Asymptomatic subjects (age range, 50 to 75 years) provided stool specimens on cards from three consecutive days for fecal occult-blood testing, which were rehydrated for interpretation. They then underwent colonoscopy. Sigmoidoscopy was defined in this study as examination of the rectum and sigmoid colon during colonoscopy, and sensitivity was estimated by determining how many patients with advanced neoplasia had an adenoma in the rectum or sigmoid colon. Advanced colonic neoplasia was defined as an adenoma 10 mm or more in diameter, a villous adenoma, an adenoma with high-grade dysplasia, or invasive cancer. Classification of subjects according to the findings was based on the most advanced lesion. **RESULTS:** A total of 2885 subjects returned the three specimen cards for fecal occult-blood testing and underwent a complete colonoscopic examination. A total of 23.9 percent of subjects with advanced neoplasia had a positive test for fecal occult blood. As compared with subjects who had a negative test for fecal occult blood, the relative risk of advanced neoplasia in subjects who had a positive test was 3.47 (95 percent confidence interval, 2.76 to 4.35). Sigmoidoscopy identified 70.3 percent of all subjects with advanced neoplasia. Combined one-time screening with a fecal occult-blood test and sigmoidoscopy identified 75.8 percent of subjects with advanced neoplasia. **CONCLUSIONS:** One-time screening with both a fecal occult-blood test with rehydration and sigmoidoscopy fails to detect advanced colonic neoplasia in 24 percent of subjects with the condition.

Question: Screening

Study Type: Prospective Cohort

Learner Level: Intermediate / Advanced

Notes: Very well done study to look at issues of screening for colon cancer with FOBTs and colonoscopy. Good paper for discussion of issues of screening. Also good paper for discussion of application to clinical practice for a screening question that would have significant cost effect if globally applied....

Lighter, J., et al. (2009). "Latent tuberculosis diagnosis in children by using the QuantiFERON-TB Gold In-Tube test." Pediatrics 123(1): 30-37.

BACKGROUND: The QuantiFERON-TB Gold test was the first blood test to be approved for the diagnosis of latent tuberculosis infection. Although it has been shown to be sensitive and specific in adults, limited data on its performance in children are available. **METHODS:** This was a prospective study of children receiving health care in New York, New York. Each child was assessed for risk factors for Mycobacterium tuberculosis infection, underwent tuberculin skin testing, and had a QuantiFERON-TB Gold In-Tube test performed. The concordance between tuberculin skin test and QuantiFERON-TB Gold In-Tube test results was calculated, and the results were analyzed according to the likelihood of exposure to M tuberculosis. **RESULTS:** Data for 207 children with valid tuberculin skin test and QuantiFERON-TB Gold In-Tube test results were analyzed. There was excellent correlation between negative tuberculin skin test results and negative QuantiFERON-TB Gold In-Tube test results; however, only 23% of children with positive tuberculin skin test results had positive QuantiFERON-TB Gold In-Tube test results. Positive QuantiFERON-TB Gold In-Tube test results were associated with increased likelihood of M tuberculosis exposure, and interferon gamma levels were higher in children with known recent exposure to M tuberculosis, compared with children with older exposure histories. Younger children produced lower interferon gamma levels in response to the mitogen (phytohemagglutinin) control used in the QuantiFERON-TB Gold In-Tube test, but indeterminate results were low for children of all ages. Performance characteristics were similar across all age groups. **CONCLUSION:** The QuantiFERON-TB Gold In-Tube test is a specific test for M tuberculosis exposure in children, with performance characteristics similar to those for adults residing in regions with low levels of endemic disease. Concerns about test sensitivity, especially for children <2 years of age, will require additional prospective long-term evaluation.

Question: Diagnosis

Study Type: Prospective Blind Comparison

Learner Level: Beginner

Notes: The study evaluates Quantiferon gold testing and is good for calculating sensitivity and specificity and defining those terms.

Maisel, A. S., et al. (2002). "Rapid measurement of B-type natriuretic peptide in the emergency diagnosis of heart failure." N Engl J Med 347(3): 161-167.

BACKGROUND: B-type natriuretic peptide is released from the cardiac ventricles in response to increased wall tension. **METHODS:** We conducted a prospective study of 1586 patients who came to the emergency department with acute dyspnea and whose B-type natriuretic peptide was measured with a bedside assay. The clinical diagnosis of congestive heart failure was adjudicated by two independent cardiologists, who were blinded to the results of the B-type natriuretic peptide assay. **RESULTS:** The final diagnosis was dyspnea due to congestive heart failure in 744 patients (47 percent), dyspnea due to noncardiac causes in 72 patients with a history of left ventricular dysfunction (5 percent), and no finding of congestive heart failure in 770 patients (49 percent). B-type natriuretic peptide levels by themselves were more accurate than any historical or physical findings or laboratory values in identifying congestive heart failure as the cause of dyspnea. The diagnostic accuracy of B-type natriuretic peptide at a cutoff of 100 pg per milliliter was 83.4 percent. The negative predictive value of B-type natriuretic peptide at levels of less than 50 pg per milliliter was 96 percent. In multiple logistic-regression analysis, measurements of B-type natriuretic peptide added significant independent

predictive power to other clinical variables in models predicting which patients had congestive heart failure. CONCLUSIONS: Used in conjunction with other clinical information, rapid measurement of B-type natriuretic peptide is useful in establishing or excluding the diagnosis of congestive heart failure in patients with acute dyspnea.

Question: Diagnosis

Study Type: Prospective

Learner Level: Beginner

Notes: great for calculating LR_s (and show that better to rule out than rule in CHF) from sensitivity/specificity data and also has nice ROC curve

McGee, S. R. and E. J. Boyko (1998). "Physical examination and chronic lower-extremity ischemia: a critical review." Arch Intern Med 158(12): 1357-1364.

OBJECTIVE: To determine the clinical utility of physical examination in patients with suspected chronic ischemia of the lower extremities. **DATA SOURCES:** MEDLINE search (January 1966 to January 1997), personal files, and bibliographies of textbooks on physical diagnosis, surgery, and vascular surgery. **STUDY SELECTION:** Both authors independently graded the studies as level 1, 2, or 3, according to predetermined criteria. Criteria deemed essential for analysis of sensitivity, specificity, and likelihood ratios were (1) clear definition of study population, (2) clear definition of physical examination maneuver, and (3) use of an acceptable criterion standard test for comparison. **RESULTS:** The following positive findings help clinicians diagnose the presence of peripheral arterial disease: abnormal pedal pulses, a unilaterally cool extremity, a prolonged venous filling time, and a femoral bruit. Other physical signs help determine the extent and distribution of vascular disease, including an abnormal femoral pulse, lower-extremity bruits, warm knees, and the Buerger test. The capillary refill test and the findings of foot discoloration, atrophic skin, and hairless extremities are unhelpful in diagnostic decisions. Mathematical formulas, derived from 2 studies using multivariate analysis, allow clinicians to estimate the probability of peripheral arterial disease in their patients. **CONCLUSION:** Certain aspects of the physical examination help clinicians make accurate judgments about the presence of peripheral arterial disease and its distribution.

Question: Diagnosis

Study Type: Systematic Review

Learner Level: Beginner / Intermediate

Notes: Solid methods with comparison to reference standards. Although this is not a meta-analysis (i.e. they didn't combine results) it is a good systematic review; Good discussion points; Good paper for discussion of diagnosis, kappa (interobserver agreement) and likelihood ratios as well as systematic review. Down side: because it is not a meta-analysis, you can't discuss certain issues such as heterogeneity

Miller, A. B., et al. (2014). "Twenty five year follow-up for breast cancer incidence and mortality of the Canadian National Breast Screening Study: randomised screening trial." BMJ 348: g366.

OBJECTIVE: To compare breast cancer incidence and mortality up to 25 years in women aged 40-59 who did or did not undergo mammography screening. **DESIGN:** Follow-up of randomised screening trial by centre coordinators, the study's central office, and linkage to cancer registries and vital statistics databases. **SETTING:** 15 screening centres in six Canadian provinces, 1980-85 (Nova Scotia, Quebec, Ontario, Manitoba, Alberta, and British Columbia). **PARTICIPANTS:** 89,835 women, aged 40-59, randomly assigned to mammography (five annual mammography screens) or control (no mammography). **INTERVENTIONS:** Women aged 40-49 in the mammography arm and all women aged 50-59 in both arms received annual physical breast examinations. Women aged 40-49 in the control arm received a single examination followed by usual care in the community. **MAIN OUTCOME MEASURE:** Deaths from breast cancer. **RESULTS:** During the five year screening period, 666 invasive breast cancers were diagnosed in the mammography arm (n=44,925 participants) and 524 in the controls (n=44,910), and of these, 180 women in the mammography arm and 171 women in the control arm died of breast cancer during the 25 year follow-up period. The overall hazard ratio for death from breast cancer diagnosed during the screening period associated with mammography was 1.05 (95% confidence interval 0.85 to 1.30). The findings for women aged 40-49 and 50-59 were almost identical. During the entire study period, 3250 women in the mammography arm and 3133 in the control arm had a diagnosis of breast cancer, and 500 and 505, respectively, died of breast cancer. Thus the cumulative mortality from breast cancer was similar between women in the mammography arm and in the control arm (hazard ratio 0.99, 95% confidence interval 0.88 to 1.12). After 15 years of follow-up a residual excess of 106 cancers was observed in the mammography arm, attributable to over-diagnosis. **CONCLUSION:** Annual mammography in women aged 40-59 does not reduce mortality from breast cancer beyond that of physical examination or usual care when adjuvant therapy for breast cancer is freely available. Overall, 22% (106/484) of screen detected invasive breast cancers were over-diagnosed, representing one over-diagnosed breast cancer for every 424 women who received mammography screening in the trial.

Question: Diagnosis

Study Type: .

Learner Level: Intermediate

Notes: Good for discussion on NNS and whether the magnitude of the impact is clinically significant (especially with all of the controversy of breast cancer screening these days)

Mittal, M. K., et al. (2012). "A clinical decision rule to identify infants with apparent life-threatening event who can be safely discharged from the emergency department." Pediatric Emergency Care 28(7): 599-605.

OBJECTIVE: This study aimed to formulate a clinical decision rule (CDR) to identify infants with apparent-life threatening event (ALTE) who are at low risk of adverse outcome and can be discharged home safely from the emergency department (ED). **METHODS:** This is a prospective cohort study of infants with an ED diagnosis of ALTE at an urban children's hospital. Admission was considered warranted if the infant required significant intervention during the hospital stay. Logistic regression and recursive partitioning were used to develop a CDR identifying patients at low risk of significant intervention and thus suitable for discharge from the ED. **RESULTS:** A total of 300 infants were enrolled; 228 (76%) were admitted; 37 (12%) required significant intervention. None died during hospital stay or within 72 hours of discharge or were diagnosed with serious bacterial infection. Logistic regression identified prematurity, abnormal result in the physical examination, color change to cyanosis, absence of symptoms of upper respiratory tract infection, and absence of choking as predictors for significant intervention. These variables were used to create a CDR, based on which, 184 infants (64%) could be discharged home safely from the ED, reducing the hospitalization rate to 102 (36%). The model has a negative predictive value of 96.2% (92%-98.3%). **CONCLUSIONS:** Only 12% of infants presenting to the ED with ALTE had a significant intervention warranting hospital admission. We created a CDR that would have decreased the admission rate safely by 40%.

Question: Clinical Decision Rule

Study Type: .

Learner Level: .

Notes: Apparent Life-Threatening Events (ALTEs): reasonable methodology. The teaching points 1) the definition of ALTE = huge catchment of ultimate diagnoses and wide variability in the spectrum of level of illness 2) outcomes measured = many were not clinically important and 3) the likelihood ratios for a positive and negative result of the CDR. The LR's were not very powerful (LR + 1.03, LR- 0.45) - the point of the study was to see if this CDR could identify young children who had an ALTE who could be safely discharged. The post-test prob even with the negative LR was that 7% of children would need a significant intervention in the hospital (baseline 12%) - most ED physicians would not feel comfortable discharging kids with this post-test prob.

Mueller, C., et al. (2004). "Use of B-type natriuretic peptide in the evaluation and management of acute dyspnea." N Engl J Med 350(7): 647-654.

BACKGROUND: B-type natriuretic peptide levels are higher in patients with congestive heart failure than in patients with dyspnea from other causes. **METHODS:** We conducted a prospective, randomized, controlled study of 452 patients who presented to the emergency department with acute dyspnea: 225 patients were randomly assigned to a diagnostic strategy involving the measurement of B-type natriuretic peptide levels with the use of a rapid bedside assay, and 227 were assessed in a standard manner. The time to discharge and the total cost of treatment were the primary end points. **RESULTS:** Base-line demographic and clinical characteristics were well matched between the two groups. The use of B-type natriuretic peptide levels reduced the need for hospitalization and intensive care; 75 percent of patients in the B-type natriuretic peptide group were hospitalized, as compared with 85 percent of patients in the control group (P=0.008), and 15 percent of those in the B-type natriuretic peptide group required intensive care, as compared with 24 percent of those in the control group (P=0.01). The median time to discharge was 8.0 days in the B-type natriuretic peptide group and 11.0 days in the control group (P=0.001). The mean total cost of treatment was 5,410 dollars (95 percent confidence interval, 4,516 dollars to 6,304 dollars) in the B-type natriuretic peptide group, as compared with 7,264 dollars (95 percent confidence interval, 6,301 dollars to 8,227 dollars) in the control group (P=0.006). The respective 30-day mortality rates were 10 percent and 12 percent (P=0.45). **CONCLUSIONS:** Used in conjunction with other clinical information, rapid measurement of B-type natriuretic peptide in the emergency department improved the evaluation and treatment of patients with acute dyspnea and thereby reduced the time to discharge and the total cost of treatment.

Question: Diagnosis

Study Type: RCT of a testing strategy

Learner Level: Intermediate

Notes: This is a randomized controlled trial in patients presenting to ER with CHF in which the intervention is a diagnostic testing strategy. The study is well-designed and measured important outcomes. ACP Journal Club Summary 2004; 141: 35. Good discussion points: Good example of new approach to diagnostic testing, randomizing to two diagnostic test strategies (using BNP vs. conventional) and looking at outcomes of hospitalization, length of stay, ICU care; Useful for discussing blinding-ie. how might the BNP result have affected clinical care in addition to diagnosing heart failure (clinicians more aggressive about COPD treatment, etc)? Can discuss how this paper might be applied to practice. Can this RCT be extrapolated to non-ER settings? Has this paper increased use of B-NP in your setting?

Nasreddine, Z. S., et al. (2005). "The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment." J Am Geriatr Soc 53(4): 695-699.

OBJECTIVES: To develop a 10-minute cognitive screening tool (Montreal Cognitive Assessment, MoCA) to assist first-line physicians in detection of mild cognitive impairment (MCI), a clinical state that often progresses to dementia. **DESIGN:** Validation study. **SETTING:** A community clinic and an academic center. **PARTICIPANTS:** Ninety-four patients meeting MCI clinical criteria supported by psychometric measures, 93 patients with mild Alzheimer's disease (AD) (Mini-Mental State Examination (MMSE) score > or =17), and 90 healthy elderly controls (NC). **MEASUREMENTS:** The MoCA and MMSE were administered to all participants, and sensitivity and specificity of both measures were assessed for detection of MCI and mild AD. **RESULTS:** Using a cutoff score 26, the MMSE had a sensitivity of 18% to detect MCI, whereas the MoCA detected 90% of MCI subjects. In the mild AD group, the MMSE had a sensitivity of 78%, whereas the MoCA detected 100%. Specificity was excellent for both MMSE and MoCA (100% and 87%, respectively). **CONCLUSION:** MCI as an entity is evolving and somewhat controversial. The MoCA is a brief cognitive screening tool with high sensitivity and specificity for detecting MCI as currently conceptualized in patients performing in the normal range on the MMSE.

Question: Diagnosis

Study Type: Case-Control

Learner Level: Beginner

Notes: challenges learners to judge the impact of weak sample selection on validity

Perrier, A., et al. (2001). "Performance of helical computed tomography in unselected outpatients with suspected pulmonary embolism." Ann Intern Med 135(2): 88-97.

BACKGROUND: Helical computed tomography (CT) is commonly used to diagnose pulmonary embolism, although its operating characteristics have been insufficiently evaluated. **OBJECTIVE:** To assess the sensitivity and specificity of helical CT in suspected pulmonary embolism. **DESIGN:** Observational study. **SETTING:** Emergency department of a teaching and community hospital. **PATIENTS:** 299 patients with clinically suspected pulmonary embolism and a plasma D-dimer level greater than 500 microgram/L. **INTERVENTION:** Pulmonary embolism was established by using a validated algorithm that included clinical assessment, lower-limb compression ultrasonography, lung scanning, and pulmonary angiography. **MEASUREMENTS:** Sensitivity, specificity, and likelihood ratios of helical CT and interobserver agreement. Helical CT scans were withheld from clinicians and were read 3 months after acquisition by radiologists blinded to all clinical data. **RESULTS:** 118 patients (39%) had pulmonary embolism. In 12 patients (4%), 2 of whom had pulmonary embolism, results of helical CT were inconclusive. For patients with conclusive results, sensitivity of helical CT was 70% (95% CI, 62% to 78%) and specificity was 91% (CI, 86% to 95%). Interobserver agreement was high (kappa = 0.823 to 0.902). The false-negative rate was lower for helical CT used after initial negative results on ultrasonography than for helical CT alone (21% vs. 30%). Use of helical CT after normal results on initial ultrasonography and nondiagnostic results on lung scanning had a false-negative rate of only 5% and a false-positive rate of only 7%. **CONCLUSION:** Helical CT should not be used alone for suspected pulmonary embolism but could replace angiography in combined strategies that include ultrasonography and lung scanning.

Question: Diagnosis

Study Type: Prospective Cohort

Learner Level: Intermediate / Advanced

Notes: Okay... This is for those of you who object to PLOPED being on this list (for VQ scans) without equal representation of a Spiral CT paper. In fact, this is an interesting paper with clear methodology that can really make you think about diagnostic testing. However, the methods are sticky with respect to the definition of the gold standard as well as the population included for the study (see figure page 89); In a more advanced group, this paper can be fun. Not to be recommended for starters, there are lots caveats. see also Rath SW. *Annals of Internal Med* 200;132(3):227-232 for systematic review on topic.

Pouplard, C., et al. (2007). "Prospective evaluation of the '4Ts' score and particle gel immunoassay specific to heparin/PF4 for the diagnosis of heparin-induced thrombocytopenia." J Thromb Haemost 5(7): 1373-1379.

BACKGROUND: Heparin-induced thrombocytopenia (HIT) is a severe disease that is often difficult to diagnose. A clinical scoring system, the '4Ts' score, has been proposed to estimate its probability before laboratory testing, and a particle gel immunoassay (H/PF4 PaGIA) has also been developed for rapid detection of HIT antibodies. **AIM:** To evaluate the performance of both methods when HIT is suspected clinically. **METHODS:** Two hundred thirteen consecutive patients were included in four centers. The probability of HIT was evaluated using the 4Ts score blind to antibody test results. HIT was confirmed only when the serotonin release assay (SRA) was positive. **RESULTS:** The risk of HIT was evaluated by the 4Ts score as low (LowR), intermediate (IR) or high (HR) in 34.7%, 60.6% and 4.7% of patients, respectively. The negative predictive value (NPV) of the 4Ts score was 100%, as the SRA was negative in all LowR patients. PaGIA was negative in 176 patients without HIT (99.4%, NPV) and the negative likelihood ratio (LR-) was 0.05. PaGIA was positive in 37 patients, including 21 with HIT (positive predictive value = 56.8%), with a positive LR of 11.4. A negative PaGIA result decreased the probability of HIT in IR patients from 10.9% before assay to 0.6%, whereas a positive result did not substantially increase the likelihood for HIT. **CONCLUSION:** The use of the 4Ts score with PaGIA appears to be a reliable strategy to rule out HIT.

Question: Diagnosis

Study Type: PropsectiveCohort

Learner Level: Beginner

Notes: Good to teach likelihood ratios and clinical prediction rules

Rathbun, S. W., et al. (2000). "Sensitivity and specificity of helical computed tomography in the diagnosis of pulmonary embolism: a systematic review." Ann Intern Med 132(3): 227-232.

PURPOSE: To determine the sensitivity and specificity of helical computed tomography (CT) for the diagnosis of pulmonary embolism and to determine the safety of withholding anticoagulant therapy in patients who have clinically suspected pulmonary embolism and negative results on helical CT. **DATA SOURCES:** The MEDLINE database was searched for all reports published from 1986 to October 1999 that evaluated the use of helical CT for the diagnosis of pulmonary embolism. Bibliographies of the retrieved articles were cross-checked to identify additional studies. **STUDY SELECTION:** All prospective English-language studies were selected. Retrospective studies, review articles, and case reports were excluded, and 5 of the 20 identified articles were excluded. The scientific validity of the remaining 15 articles was assessed. **DATA EXTRACTION:** Two of the authors used a priori, pre-defined criteria to independently assess each study. A third author resolved disagreements by adjudication. The pre-defined criteria were inclusion of a consecutive series of all patients with suspected pulmonary embolism, inclusion of patients with and those without pulmonary embolism, a broad spectrum of patient characteristics, performance of helical CT and pulmonary angiography (or an appropriate reference test) in all patients, and independent interpretation of the CT scan and pulmonary angiogram (or reference test). Specific data on sensitivity and specificity and the associated 95% CIs were recorded when available. **DATA SYNTHESIS:** No study met all of the predefined criteria for adequately evaluating sensitivity and specificity. The reported sensitivity of helical CT ranged from 53% to 100%, and specificity ranged from 81% to 100%. In no prospective study was anticoagulant therapy withheld without further testing for venous thromboembolism in consecutive patients with suspected pulmonary embolism. One prospective study reported the outcome of selected patients with negative results on helical CT who did not receive anticoagulant therapy. **CONCLUSIONS:** Use of helical CT in the diagnosis of pulmonary embolism has not been adequately evaluated. The safety of withholding anticoagulant treatment in patients with negative results on helical CT is uncertain. Definitive large, prospective studies should be done to evaluate the sensitivity, specificity, and safety of helical CT for diagnosis of suspected pulmonary embolism.

Question: Diagnosis

Study Type: Systematic Review

Learner Level: Intermediate / Advanced

Notes: Important paper that brought together the evidence regarding helical CT. Good discussion points: Good for discussion of how papers are graded in a systematic review. Note: this paper predates the other diagnostic paper about helical CT in the diagnosis section. Interesting to consider whether the two papers come to similar conclusions. See also Perrier A. *Annals of Internal Medicine* 2001;125(2):88-97 for example of an original study

Rothman, R., et al. (2003). "Does this child have acute otitis media?" JAMA 290(12): 1633-1640.

CONTEXT: Acute otitis media (AOM) is one of the most common problems in pediatrics. An accurate diagnosis of AOM can guide proper treatment and follow-up. **OBJECTIVE:** To systematically review the literature regarding precision and accuracy of history taking and physical examination in diagnosing AOM in children. **DATA SOURCES:** We searched MEDLINE for English-language articles published from 1966 through May 2002. Bibliographies of retrieved articles and textbooks were also searched. **STUDY SELECTION:** We located studies with original data on the precision or accuracy of history or physical examination for AOM in children. Of 397 references initially identified, 6 met inclusion criteria for analysis. **DATA EXTRACTION:** Two authors independently reviewed and abstracted data to calculate likelihood ratios (LRs) for symptoms and signs. **DATA SYNTHESIS:** Four studies of symptoms used clinical diagnosis as the criterion standard and were limited by incorporation bias. Ear pain is the most useful symptom (positive LRs, 3.0-7.3); fever, upper respiratory tract symptoms, and irritability are less useful. One study of clinical signs used tympanocentesis as the criterion standard, and we adjusted the results to correct for verification bias. A cloudy (adjusted LR, 34; 95% confidence interval [CI], 28-42), bulging (adjusted LR, 51; 95% CI, 36-73), or distinctly immobile (adjusted LR, 31; 95% CI, 26-37) tympanic membrane on pneumatic otoscopy are the most useful signs for detecting AOM. A distinctly red tympanic membrane is also helpful (adjusted LR, 8.4; 95% CI, 6.7-11) whereas a normal color makes AOM much less likely (adjusted LR, 0.2; 95% CI, 0.19-0.21). **CONCLUSIONS:** Although many of the studies included in this analysis are limited by bias, a cloudy, bulging, or clearly immobile tympanic membrane is most helpful for detecting AOM. The degree of erythema may also be useful since a normal color makes otitis media unlikely whereas a distinctly red tympanic membrane increases the likelihood significantly.

Question: Diagnosis

Study Type: Meta-analysis

Learner Level: Intermediate / Advanced

Notes: Another meta-analysis is part of the Rational Clinical Exam Series. This article on otitis media in children is relevant to anyone who has ever had (or ever will have) a cranky child with a fever...Good discussion points: As with all Rational Clinical Exam articles, this is an evidence summary of diagnostic tests- thus one can discuss both systematic review methodology AND diagnosis, specifically Likelihood ratio. Excellent paragraph under statistical methods (Page 990) that defines likelihood ratio as well as diagnostic odds ratio. Stumbling block may be the large number of items listed in Tables 2 and 3 (p 992 and 993), however a clear difference in data can be noted in patients older than 60 years compared to all comers.

Schneider, H. G., et al. (2009). "B-type natriuretic peptide testing, clinical outcomes, and health services use in emergency department patients with dyspnea: a randomized trial." Ann Intern Med 150(6): 365-371.

BACKGROUND: B-type natriuretic peptide (BNP) is used to diagnose heart failure, but the effects of using the test on all dyspneic patients is uncertain. **OBJECTIVE:** To assess whether BNP testing alters clinical outcomes and health services use of acutely dyspneic patients. **DESIGN:** Randomized, single-blind study. Patients were assigned to a treatment group through randomized numbers in a sealed envelope. Patients were blinded to the intervention, but clinicians and those who assessed trial outcomes were not. **SETTING:** 2 Australian teaching hospital emergency departments. **PATIENTS:** 612 consecutive patients who presented with acute severe dyspnea from August 2005 to March 2007. **INTERVENTION:** BNP testing (n = 306) or no testing (n = 306). **MEASUREMENTS:** Admission rates, length of stay, and emergency department medications (primary outcomes); mortality and readmission rates (secondary outcomes). **RESULTS:** There were no between-group differences in hospital admission rates (85.6% [BNP group] vs. 86.6% [control group]; difference, -1.0 percentage point [95% CI, -6.5 to 4.5 percentage points]; P = 0.73), length of admission (median, 4.4 days [interquartile range, 2 to 9 days] vs. 5.0 days [interquartile range, 2 to 9 days]; P = 0.94), or management of patients in the emergency department. Test discrimination was good (area under the receiver-operating characteristic curve, 0.87 [CI, 0.83 to 0.91]). Adverse events were not measured. **LIMITATION:** Most patients were very short of breath and required hospitalization; the findings might not apply for evaluating patients with milder degrees of breathlessness. **CONCLUSION:** Measurement of BNP in all emergency department patients with severe shortness of breath had no apparent effects on clinical outcomes or use of health services. The findings do not support routine use of BNP testing in all severely dyspneic patients in the emergency department. **PRIMARY FUNDING SOURCE:** Janssen-Cilag.

Question: Diagnosis

Study Type: .

Learner Level: Intermediate

Notes: This is a diagnostic testing article in which a strategy for testing is tested using a randomized design. Some learners may get confused about the interface between diagnosis and 'treatment'—as you need to use an RCT critical appraisal sheet (usually for therapy and RCT), not a diagnosis sheet (usually for determining test characteristics). This is a good paper for teaching how to determine the impact of testing strategies on care. (This paper has an ACP Journal Club summary)

Sivitz, A. B., et al. (2014). "Evaluation of acute appendicitis by pediatric emergency physician sonography." Ann Emerg Med 64(4): 358-364.e354.

STUDY OBJECTIVE: We investigate the accuracy of pediatric emergency physician sonography for acute appendicitis in children. **METHODS:** We prospectively enrolled children requiring surgical or radiology consultation for suspected acute appendicitis at an urban pediatric emergency department. Pediatric emergency physicians performed focused right lower-quadrant sonography after didactics and hands-on training with a structured scanning algorithm, including the graded-compression technique. We compared their sonographic interpretations with clinical and radiologic findings, as well as clinical outcomes as defined by follow-up or pathologic findings. **RESULTS:** Thirteen pediatric emergency medicine sonographers performed 264 ultrasonographic studies, including 85 (32%) in children with pathology-verified appendicitis. Bedside sonography had a sensitivity of 85% (95% confidence interval [CI] 75% to 95%), specificity of 93% (95% CI 85% to 100%), positive likelihood ratio of 11.7 (95% CI 6.9 to 20), and negative likelihood ratio of 0.17 (95% CI 0.1 to 0.28). **CONCLUSION:** With focused ultrasonographic training, pediatric emergency physicians can diagnose acute appendicitis with substantial accuracy.

Question: .

Study Type: .

Learner Level: .

Notes: Use of bedside ultrasound for assessment of pediatric appendicitis: Nice teaching article in that it's pretty straight forward with regard to the validity criteria. Relatively easy from the Figure to calculate the test characteristics. It's a timely article in 2016 as ultrasound use is increasing significantly outside of the realm of radiologists.

Swensen, S. J., et al. (1997). "The probability of malignancy in solitary pulmonary nodules. Application to small radiologically indeterminate nodules." Arch Intern Med 157(8): 849-855.

BACKGROUND: A clinical prediction model to identify malignant nodules based on clinical data and radiological characteristics of lung nodules was derived using logistic regression from a random sample of patients (n = 419) and tested on data from a separate group of patients (n = 210). **OBJECTIVE:** To use multivariate logistic regression to estimate the probability of malignancy in radiologically indeterminate solitary pulmonary nodules (SPNs) in a clinically relevant subset of patients with SPNs that measured between 4 and 30 mm in diameter. **PATIENTS AND METHODS:** A retrospective cohort study at a multispecialty group practice included 629 patients (320 men, 309 women) with newly discovered (between January 1, 1984, and May 1, 1986) 4- to 30-mm radiologically indeterminate SPNs on chest radiography. Patients with a diagnosis of cancer within 5 years prior to the discovery of the nodule were excluded. Clinical data included age, sex, cigarette-smoking status, and history of extrathoracic malignant neoplasm, asbestos exposure, and chronic interstitial or obstructive lung disease; chest radiological data included the diameter, location, edge characteristics (eg, lobulation, spiculation, and shaginess), and other characteristics (eg, cavitation) of the SPNs. Predictors were identified in a random sample of two thirds of the patients and tested in the remaining one third. **RESULTS:** Sixty-five percent of the nodules were benign, 23% were malignant, and 12% were indeterminate. Three clinical characteristics (age,

cigarette-smoking status, and history of cancer [diagnosis, > or = 5 years ago]) and 3 radiological characteristics (diameter, spiculation, and upper lobe location of the SPNs) were independent predictors of malignancy. The area (+/-SE) under the evaluated receiver operating characteristic curve was 0.8328 +/- 0.0226. CONCLUSION: Three clinical and 3 radiographic characteristics predicted the malignancy in radiologically indeterminate SPNs.

Question: Diagnosis

Study Type: Clinical Prediction Rule

Learner Level: Advanced

Notes: Derivation and validation of Clinical Prediction Rules for diagnosis.

Taylor, R. A., et al. (2016). "Determination of a Testing Threshold for Lumbar Puncture in the Diagnosis of Subarachnoid Hemorrhage after a Negative Head Computed Tomography: A Decision Analysis." *Acad Emerg Med* 23(10): 1119-1127.

OBJECTIVE: The objective was to determine the testing threshold for lumbar puncture (LP) in the evaluation of aneurysmal subarachnoid hemorrhage (SAH) after a negative head computed tomography (CT). As a secondary aim we sought to identify clinical variables that have the greatest impact on this threshold. **METHODS:** A decision analytic model was developed to estimate the testing threshold for patients with normal neurologic findings, being evaluated for SAH, after a negative CT of the head. The testing threshold was calculated as the pretest probability of disease where the two strategies (LP or no LP) are balanced in terms of quality-adjusted life-years. Two-way and probabilistic sensitivity analyses (PSAs) were performed. **RESULTS:** For the base-case scenario the testing threshold for performing an LP after negative head CT was 4.3%. Results for the two-way sensitivity analyses demonstrated that the test threshold ranged from 1.9% to 15.6%, dominated by the uncertainty in the probability of death from initial missed SAH. In the PSA the mean testing threshold was 4.3% (95% confidence interval = 1.4% to 9.3%). Other significant variables in the model included probability of aneurysmal versus nonaneurysmal SAH after negative head CT, probability of long-term morbidity from initial missed SAH, and probability of renal failure from contrast-induced nephropathy. **CONCLUSIONS:** Our decision analysis results suggest a testing threshold for LP after negative CT to be approximately 4.3%, with a range of 1.4% to 9.3% on robust PSA. In light of these data, and considering the low probability of aneurysmal SAH after a negative CT, classical teaching and current guidelines addressing testing for SAH should be revisited.

Question: Diagnosis

Study Type: Decision Analysis

Learner Level: Intermediate/Advanced

Notes: A study of a clinical decision analysis, particularly in the approach to patients with suspected subarachnoid hemorrhage – this is a recent example of the use of clinical decision analysis for formal threshold analysis, and highlights how the careful use of evidence-based CDA, particularly the sensitivity analysis, can be useful in questioning the recommendations from current practice guidelines.

Tunariu, N., et al. (2007). "Ventilation-perfusion scintigraphy is more sensitive than multidetector CTPA in detecting chronic thromboembolic pulmonary disease as a treatable cause of pulmonary hypertension." *Journal of Nuclear Medicine* 48(5): 680-684.

Pulmonary hypertension (PH) is a progressive disease with a poor prognosis. Identifying chronic thromboembolic pulmonary disease as a cause of PH has major clinical implications as these patients could be potentially offered a surgical cure. Ventilation-perfusion (V/Q) scintigraphy has a high sensitivity to detect embolic disease but its value has been challenged with the emergence of multidetector CT pulmonary angiography (CTPA). We compared the value of V/Q scintigraphy with CTPA in detecting chronic thromboembolic pulmonary disease. **METHODS:** We retrospectively reviewed the results of V/Q scintigraphy and CTPA performed on patients who had been referred to the Pulmonary Hypertension Service at Hammersmith Hospital between 2000 and 2005. A total of 227 patients (85 males, 142 females; age range, 18-81 y; mean age, 42 y) had all tests done at Hammersmith Hospital and were included in the study. Interpretation of scans was according to the modified PIOPED (Prospective Investigation of Pulmonary Embolism Diagnosis) criteria. CTPA was considered as suggestive of chronic thromboembolic pulmonary disease if it showed visualization of the thrombus or webs, recanalization, perfusion abnormalities, stenosis, or strictures. Standard pulmonary angiography was performed via femoral approach. In 90% of the cases, CTPA and V/Q scintigraphy were performed within 10 d. **RESULTS:** Seventy-eight patients (group A) had a final diagnosis of chronic thromboembolic pulmonary hypertension (CTEPH) and 149 (group B) had non-CTEPH etiology. Among group A, V/Q scintigraphy was reported as high probability in 75 patients, intermediate probability in 1 patient, and low probability in 2 patients. CTPA was positive in 40 patients and negative in 38 patients. Among group B, V/Q scintigraphy was reported as low probability in 134, intermediate probability in 7, and high probability in 8 patients. CTPA was negative in 148 patients and false-positive in 1 patient. Statistical analysis showed V/Q scintigraphy to have a sensitivity of 96%-97.4% and a specificity of 90%-95%. CTPA showed a sensitivity of 51% and a specificity of 99%. **CONCLUSION:** Our results demonstrate that V/Q scintigraphy has a higher sensitivity than CTPA in detecting CTEPH as a potential curable cause of PH.

Question: Diagnosis

Study Type: Retrospective

Learner Level: .

Notes: Diagnosis

Tuon, F. F., et al. (2006). "Adenosine deaminase and tuberculous pericarditis--a systematic review with meta-analysis." Acta Trop 99(1): 67-74.

BACKGROUND: Adenosine deaminase (ADA) activity in pericardial fluid is a valuable aid in the diagnosis of tuberculous pericarditis (TP), but there is no systematic review performed to evaluate the benefits of ADA activity as an adjunctive test for TP diagnosis. The objective of this systematic review was to evaluate the utility of ADA activity as a diagnostic marker of TP on patients presenting with pericardial effusion. **METHODS:** MEDLINE, LILACS and Cochrane Library databases (1980-2005) searches to identify articles related to adenosine deaminase activity on TP diagnosis. Articles with patients with at least one TP diagnostic criteria were included. The controls were patients with other pericardial diseases with moderate or large pericardial effusion. To calculate the sensitivity, specificity, as well as positive and negative likelihood ratios we extracted the total number of confirmed TP cases over all patients with pericardial effusion as well as the number of cases with ADA activity values of 40 U/L and over. **RESULTS:** Thirty one studies met our initial inclusion criteria and five articles were selected. The heterogeneity limited the specificity analysis ($p=0.004$). The method yielded a sensitivity and specificity of 88% and 83%, respectively. The SROC curve presented an area with a tendency towards 1 (value of 0.9539) and corroborates the diagnostic value of ADA activity. **CONCLUSIONS:** The present study confirms the clinical value of ADA activity as adjunctive diagnostic marker of TP among other causes of pericardial effusion.

Question: Diagnosis

Study Type: Meta-analysis

Learner Level: Advanced

Notes: Good to teach likelihood ratios

Tzimenatos, L., et al. (2018). "Accuracy of the Urinalysis for Urinary Tract Infections in Febrile Infants 60 Days and Younger." Pediatrics 141(2).

OBJECTIVES: Reports of the test accuracy of the urinalysis for diagnosing urinary tract infections (UTIs) in young febrile infants have been variable. We evaluated the test characteristics of the urinalysis for diagnosing UTIs, with and without associated bacteremia, in young febrile infants. **METHODS:** We performed a planned secondary analysis of data from a prospective study of febrile infants ≤ 60 days old at 26 emergency departments in the Pediatric Emergency Care Applied Research Network. We evaluated the test characteristics of the urinalysis for diagnosing UTIs, with and without associated bacteremia, by using 2 definitions of UTI: growth of $\geq 50\,000$ or $\geq 10\,000$ colony-forming units (CFUs) per mL of a uropathogen. We defined a positive urinalysis by the presence of any leukocyte esterase, nitrite, or pyuria (>5 white blood cells per high-power field). **RESULTS:** Of 4147 infants analyzed, 289 (7.0%) had UTIs with colony counts $\geq 50\,000$ CFUs/mL, including 27 (9.3%) with bacteremia. For these UTIs, a positive urinalysis exhibited sensitivities of 0.94 (95% confidence interval [CI]: 0.91-0.97), regardless of bacteremia; 1.00 (95% CI: 0.87-1.00) with bacteremia; and 0.94 (95% CI: 0.90-0.96) without bacteremia. Specificity was 0.91 (95% CI: 0.90-0.91) in all groups. For UTIs with colony counts $\geq 10\,000$ CFUs/mL, the sensitivity of the urinalysis was 0.87 (95% CI: 0.83-0.90), and specificity was 0.91 (95% CI: 0.90-0.92). **CONCLUSIONS:** The urinalysis is highly sensitive and specific for diagnosing UTIs, especially with $\geq 50\,000$ CFUs/mL, in febrile infants ≤ 60 days old, and particularly for UTIs with associated bacteremia.

Question: Diagnosis

Study Type: .

Learner Level: Intermediate

Notes: Large multi-centered study. Planned secondary analysis of large prospective cohort. Can talk about study design – is this a prospective study or a retrospective study? Can have a good discussion on the inclusion and exclusion criteria – are these reasonable? Is this what you would do when designing a study to answer this question? If not, what would you do differently. Reports LRs!! Test characteristics are reported for individual aspects of urinalysis (Leuk esterase, NO₂, WBC) as well as aggregate positive or negative UA. Is this OK? Does this help you when interpreting the UA at the bedside? LRs (both + and -) are substantial enough that the significantly affect your pre- to post-test probabilities.

Harm

Autier, P., et al. (1997). "Melanoma risk and residence in sunny areas. EORTC Melanoma Co-operative Group. European Organization for Research and Treatment of Cancer." British Journal of Cancer 76(11): 1521-1524.

Melanoma risk among subjects from Germany, France and Belgium who had lived for 1 year or more in sunny climates was examined in a one-to-one unmatched case-control study conducted among white subjects 20 years old or more. A total of 412 consecutive patients with melanoma diagnosed from 1 January 1991 onwards, were derived from hospital registers; 445 controls were randomly chosen in the same municipality as the cases. After adjustment for host characteristics, melanoma risk associated with residence in a sunny area was 2.7 (95% CI: 1.4-5.2), increasing to 4.7 (95% CI: 1.4-13.5) if subjects sought a suntan when residing in sunny climates, and to 4.3 (95% CI: 1.7-11.1) if subjects arrived before the age of 10 years in the sunny area. Residence in sunny areas and recreational sun exposure seemed to combine their effects on melanoma risk. Increase in melanoma risk conveyed by deliberate sun exposure during adulthood was highest among subjects who had lived in sunny areas as a child or adolescent and lowest among subjects who had never resided in sunny areas. Our results support conclusions from migrant studies that indicated that childhood is a critical period of either vulnerability to solar radiation or more frequent exposures to melanoma risk factors. They also suggest that moderate sun exposure of an adult who was heavily sun exposed in childhood is associated with a higher melanoma risk than that of high sun exposure of an adult who was sun protected in childhood.

Question: Harm

Study Type: Case-Control

Learner Level: .

Notes: Harm

Batty, G. D., et al. (2010). "Body mass index and attempted suicide: Cohort study of 1,133,019 Swedish men." American Journal of Epidemiology 172(8): 890-899.

Associations between body mass index (BMI) and attempted (nonfatal) suicide have recently been reported. However, the few existing studies are relatively small in scale, the majority cross-sectional, and results contradictory. The authors have explored BMI-attempted suicide associations in a large cohort of 1,133,019 Swedish men born between 1950 and 1976, with BMI measured in early adulthood. During a mean follow-up of 23.9 years, a total of 18,277 (1.6%) men had at least 1 hospital admission for attempted suicide. After adjustment for confounding factors, there was a stepwise, linear decrease in attempted suicide with increasing BMI across the full BMI range (per standard deviation increase in BMI, hazard ratio = 0.93, 95% confidence interval: 0.91, 0.94). Analyses excluding men with depression at baseline were essentially identical to those based on the complete cohort. In men free from depression at baseline, controlling for subsequent depression slightly attenuated the raised risk of attempted suicide, particularly in lower weight men. This study suggests that lower weight men have an increased risk of attempted suicide and that associations may extend into the "normal" BMI range.

Question: Harm

Study Type: Cohort Study

Learner Level: Beginner/Intermediate

Notes: This paper is great because the authors find a puzzling, counter-intuitive linear relationship between lower BMI and increased hospitalization for suicide attempts in this population-based cohort that begs for an explanation! Learners have the opportunity to appraise the paper and systematically examine alternative explanations for the observed association----was it spurious? did it occur by chance? was there bias? another variable associated with both that explains the association? This paper is best paired with this required reading: Chapter 9 "Enhancing Causal Inference in Observational Studies" in Designing Clinical Research by Stephen Hulley

Butt, D. A., et al. (2013). "The risk of falls on initiation of antihypertensive drugs in the elderly." Osteoporosis International 24(10): 2649-2657.

SUMMARY: Antihypertensive drugs are associated with an immediate increased falls risk in elderly patients which was significant during the first 14 days after receiving a thiazide diuretic, angiotensin-converting enzyme inhibitor, angiotensin II receptor blocker, calcium channel blocker, or beta-adrenergic blocker. Fall prevention strategies during this period may prevent fall-related injuries. **INTRODUCTION:** The purpose of this study is to evaluate if initiation of the common antihypertensive drugs is associated with the occurrence of falls. **METHODS:** This population-based self-controlled case series study used healthcare administrative databases to identify new users of antihypertensive drugs in the elderly aged 66 and older living in Ontario, Canada who suffered a fall from April 1, 2000 to March 31, 2009. The risk period was the first 45 days following antihypertensive therapy initiation, further subdivided into 0-14 and 15-44 days with control periods before and after treatment in a 450-day observation period. We calculated the relative incidence (incidence rate ratio, IRR), defined as the rate of falls in the risk period compared to falls rate in the control periods. **RESULTS:** Of the 543,572 new users of antihypertensive drugs among community-dwelling elderly, 8,893 experienced an injurious fall that required hospital care during the observation period. New users had a 69 % increased risk of having an injurious fall during the first 45 days following antihypertensive treatment (IRR = 1.69; 95 % CI, 1.57-1.81). This finding was consistent for thiazide diuretics, angiotensin-converting enzyme inhibitors, calcium channel blockers, and beta-adrenergic blockers but not angiotensin II receptor antagonists. There was also an increased falls risk during the first 14 days of antihypertensive drug initiation (IRR = 1.94; 95

% CI, 1.75-2.16), which was consistent for all antihypertensive drug classes. **CONCLUSIONS:** This study suggests that initiation of antihypertensive drugs is a risk factor for falls in the elderly. Fall prevention strategies during this period may reduce injuries.

Question: Harm

Study Type: Case Series

Learner Level: Advanced

Notes: Prognosis/Harm – more challenging but interesting article for advanced EBMers on risk of falls at various time periods following initiation of antihypertensive medications, using self-controlled case series design so each participant serves as their own control.

Castellsague, X., et al. (2002). "Male circumcision, penile human papillomavirus infection, and cervical cancer in female partners." N Engl J Med 346(15): 1105-1112.

BACKGROUND: It is uncertain whether male circumcision reduces the risks of penile human papillomavirus (HPV) infection in the man and of cervical cancer in his female partner. **METHODS:** We pooled data on 1913 couples enrolled in one of seven case-control studies of cervical carcinoma in situ and cervical cancer in five countries. Circumcision status was self-reported, and the accuracy of the data was confirmed by physical examination at three study sites. The presence or absence of penile HPV DNA was assessed by a polymerase-chain-reaction assay in 1520 men and yielded a valid result in the case of 1139 men (74.9 percent). **RESULTS:** Penile HPV was detected in 166 of the 847 uncircumcised men (19.6 percent) and in 16 of the 292 circumcised men (5.5 percent). After adjustment for age at first intercourse, lifetime number of sexual partners, and other potential confounders, circumcised men were less likely than uncircumcised men to have HPV infection (odds ratio, 0.37; 95 percent confidence interval, 0.16 to 0.85). Monogamous women whose male partners had six or more sexual partners and were circumcised had a lower risk of cervical cancer than women whose partners were uncircumcised (adjusted odds ratio, 0.42; 95 percent confidence interval, 0.23 to 0.79). Results were similar in the subgroup of men in whom circumcision was confirmed by medical examination. **CONCLUSIONS:** Male circumcision is associated with a reduced risk of penile HPV infection and, in the case of men with a history of multiple sexual partners, a reduced risk of cervical cancer in their current female partners.

Question: Harm

Study Type: Case Control (pooled)

Learner Level: Intermediate / Advanced

Notes: Pros: Allows for discussion of case definition and identification and appropriate controls. Adjustment for known risk factors. Cons: Logistic regression used to generate ORs (more difficult to discuss compared to 2x2 table). Few people have much experience critically appraising SR of Case controlled studies and new learners may be overwhelmed

Cumming, R. G., et al. (1997). "Use of inhaled corticosteroids and the risk of cataracts." N Engl J Med 337(1): 8-14.

BACKGROUND: The use of systemic corticosteroids is a risk factor for the development of posterior subcapsular cataracts, but the association between inhaled corticosteroids and cataracts is uncertain. **METHODS:** We conducted a population-based, cross-sectional study of vision and common eye diseases in an urban area of the Blue Mountains, near Sydney, Australia. We recruited 3654 people 49 to 97 years of age; the participation rate was 82 percent. We collected information by questionnaire on potential risk factors for cataracts, including the current or prior use of inhaled corticosteroids (beclomethasone or budesonide). Photographs of the subjects' lenses were graded, without information on the subjects, to determine the presence and severity of cortical, nuclear, and posterior subcapsular cataracts. **RESULTS:** Three hundred seventy subjects reported using inhaled corticosteroids, 164 currently and 206 previously. Among these subjects, after adjustment for age and sex, there was a higher prevalence of nuclear cataracts (relative prevalence, 1.5; 95 percent confidence interval, 1.2 to 1.9) and posterior subcapsular cataracts (relative prevalence, 1.9; 95 percent confidence interval, 1.3 to 2.8) than among the subjects with no inhaled-corticosteroid use, but the prevalence of cortical cataracts was not significantly higher (relative prevalence, 1.1; 95 percent confidence interval, 0.9 to 1.3). Higher cumulative lifetime doses of beclomethasone were associated with higher risks of posterior subcapsular cataracts (P for trend <0.001); the highest prevalence (27 percent) was found in subjects whose lifetime dose was over 2000 mg (relative prevalence, 5.5). Adjusting for the use of systemic corticosteroids and other potential confounders had little effect on the magnitude of the associations. The associations with posterior subcapsular cataracts, but not those with nuclear cataracts, were less marked when the analyses were restricted to subjects who had never used systemic corticosteroids. **CONCLUSIONS:** The use of inhaled corticosteroids is associated with the development of posterior subcapsular and nuclear cataracts.

Question: Harm

Study Type: Cross sectional

Learner Level: Beginner / Intermediate

Notes: This is a population based cross sectional study. Very good example of a population based sampling frame. Good discussion points: Good for discussion of prevalence, ratios and confidence intervals. Can use to discuss questions of etiology and dose response - relationships

Freedman, N. D., et al. (2012). "Association of coffee drinking with total and cause-specific mortality." *N Engl J Med* 366(20): 1891-1904.

BACKGROUND: Coffee is one of the most widely consumed beverages, but the association between coffee consumption and the risk of death remains unclear. **METHODS:** We examined the association of coffee drinking with subsequent total and cause-specific mortality among 229,119 men and 173,141 women in the National Institutes of Health-AARP Diet and Health Study who were 50 to 71 years of age at baseline. Participants with cancer, heart disease, and stroke were excluded. Coffee consumption was assessed once at baseline. **RESULTS:** During 5,148,760 person-years of follow-up between 1995 and 2008, a total of 33,731 men and 18,784 women died. In age-adjusted models, the risk of death was increased among coffee drinkers. However, coffee drinkers were also more likely to smoke, and, after adjustment for tobacco-smoking status and other potential confounders, there was a significant inverse association between coffee consumption and mortality. Adjusted hazard ratios for death among men who drank coffee as compared with those who did not were as follows: 0.99 (95% confidence interval [CI], 0.95 to 1.04) for drinking less than 1 cup per day, 0.94 (95% CI, 0.90 to 0.99) for 1 cup, 0.90 (95% CI, 0.86 to 0.93) for 2 or 3 cups, 0.88 (95% CI, 0.84 to 0.93) for 4 or 5 cups, and 0.90 (95% CI, 0.85 to 0.96) for 6 or more cups of coffee per day ($P < 0.001$ for trend); the respective hazard ratios among women were 1.01 (95% CI, 0.96 to 1.07), 0.95 (95% CI, 0.90 to 1.01), 0.87 (95% CI, 0.83 to 0.92), 0.84 (95% CI, 0.79 to 0.90), and 0.85 (95% CI, 0.78 to 0.93) ($P < 0.001$ for trend). Inverse associations were observed for deaths due to heart disease, respiratory disease, stroke, injuries and accidents, diabetes, and infections, but not for deaths due to cancer. Results were similar in subgroups, including persons who had never smoked and persons who reported very good to excellent health at baseline. **CONCLUSIONS:** In this large prospective study, coffee consumption was inversely associated with total and cause-specific mortality. Whether this was a causal or associational finding cannot be determined from our data. (Funded by the Intramural Research Program of the National Institutes of Health, National Cancer Institute, Division of Cancer Epidemiology and Genetics.)

Question: Harm

Study Type: Cohort Study

Learner Level: Beginner

Notes: cohort study on coffee consumption and mortality; good for teaching association/causation, adjusting for confounders (smoking is a major confounder) and also "dose response" of coffee and its impact on mortality, and it's a fun topic that should reassure the coffee drinkers in the room!

Frei, P., et al. (2011). "Use of mobile phones and risk of brain tumours: update of Danish cohort study." *BMJ* 343: d6387.

OBJECTIVE: To investigate the risk of tumours in the central nervous system among Danish mobile phone subscribers. **DESIGN:** Nationwide cohort study. **SETTING:** Denmark. **PARTICIPANTS:** All Danes aged ≥ 30 and born in Denmark after 1925, subdivided into subscribers and non-subscribers of mobile phones before 1995. **MAIN OUTCOME MEASURES:** Risk of tumours of the central nervous system, identified from the complete Danish Cancer Register. Sex specific incidence rate ratios estimated with log linear Poisson regression models adjusted for age, calendar period, education, and disposable income. **RESULTS:** 358,403 subscription holders accrued 3.8 million person years. In the follow-up period 1990-2007, there were 10,729 cases of tumours of the central nervous system. The risk of such tumours was close to unity for both men and women. When restricted to individuals with the longest mobile phone use—that is, ≥ 13 years of subscription—the incidence rate ratio was 1.03 (95% confidence interval 0.83 to 1.27) in men and 0.91 (0.41 to 2.04) in women. Among those with subscriptions of ≥ 10 years, ratios were 1.04 (0.85 to 1.26) in men and 1.04 (0.56 to 1.95) in women for glioma and 0.90 (0.57 to 1.42) in men and 0.93 (0.46 to 1.87) in women for meningioma. There was no indication of dose-response relation either by years since first subscription for a mobile phone or by anatomical location of the tumour—that is, in regions of the brain closest to where the handset is usually held to the head. **CONCLUSIONS:** In this update of a large nationwide cohort study of mobile phone use, there were no increased risks of tumours of the central nervous system, providing little evidence for a causal association.

Question: Harm

Study Type: Cohort Study

Learner Level: Beginner

Notes: cohort study from Denmark (land of amazing population based data) showing a lack of association between cell phone use and brain tumors

Juni, P., et al. (2004). "Risk of cardiovascular events and rofecoxib: cumulative meta-analysis." *Lancet* 364(9450): 2021-2029.

BACKGROUND: The cyclo-oxygenase 2 inhibitor rofecoxib was recently withdrawn because of cardiovascular adverse effects. An increased risk of myocardial infarction had been observed in 2000 in the Vioxx Gastrointestinal Outcomes Research study (VIGOR), but was attributed to cardioprotection of naproxen rather than a cardiotoxic effect of rofecoxib. We used standard and cumulative random-effects meta-analyses of randomised controlled trials and observational studies to establish whether robust evidence on the adverse effects of rofecoxib was available before September, 2004. **METHODS:** We searched bibliographic databases and relevant files of the US Food and Drug Administration. We included all randomised controlled trials in patients with chronic musculoskeletal disorders that compared rofecoxib with other non-steroidal anti-inflammatory drugs (NSAIDs) or placebo, and cohort and case-control studies of cardiovascular risk and naproxen. Myocardial infarction was the primary endpoint. **FINDINGS:** We identified 18 randomised controlled trials and 11 observational studies. By the end of 2000 (52 myocardial infarctions, 20742 patients) the relative risk from randomised controlled trials was 2.30 (95% CI 1.22-4.33, $p = 0.010$), and 1 year later (64 events, 21432 patients) it was 2.24 (1.24-4.02, $p = 0.007$). There was little evidence that the relative risk differed depending on the control group (placebo, non-

naproxen NSAID, or naproxen; $p=0.41$) or trial duration ($p=0.82$). In observational studies, the cardioprotective effect of naproxen was small (combined estimate 0.86 [95% CI 0.75-0.99]) and could not have explained the findings of the VIGOR trial.

INTERPRETATION: Our findings indicate that rofecoxib should have been withdrawn several years earlier. The reasons why manufacturer and drug licensing authorities did not continuously monitor and summarise the accumulating evidence need to be clarified.

Question: Therapy (Harm)

Study Type: Meta-analysis

Learner Level: Intermediate

Notes: Topical question of broad clinical interest which learners find interesting. Robust cumulative meta-analysis showing the chronology of relevant RCTs on the topic demonstrating harm of this commonly prescribed medication.

Kernan, W. N., et al. (2000). "Phenylpropanolamine and the risk of hemorrhagic stroke." *N Engl J Med* 343(25): 1826-1832.

BACKGROUND: Phenylpropanolamine is commonly found in appetite suppressants and cough or cold remedies. Case reports have linked the use of products containing phenylpropanolamine to hemorrhagic stroke, often after the first use of these products. To study the association, we designed a case-control study. **METHODS:** Men and women 18 to 49 years of age were recruited from 43 U.S. hospitals. Eligibility criteria included the occurrence of a subarachnoid or intracerebral hemorrhage within 30 days before enrollment and the absence of a previously diagnosed brain lesion. Random-digit dialing identified two matched control subjects per patient. **RESULTS:** There were 702 patients and 1376 control subjects. For women, the adjusted odds ratio was 16.58 (95 percent confidence interval, 1.51 to 182.21; $P=0.02$) for the association between the use of appetite suppressants containing phenylpropanolamine and the risk of a hemorrhagic stroke and 3.13 (95 percent confidence interval, 0.86 to 11.46; $P=0.08$) for the association with the first use of a product containing phenylpropanolamine. All first uses of phenylpropanolamine involved cough or cold remedies. For men and women combined, the adjusted odds ratio was 1.49 (95 percent confidence interval, 0.84 to 2.64; $P=0.17$) for the association between the use of a product containing phenylpropanolamine and the risk of a hemorrhagic stroke, 1.23 (95 percent confidence interval, 0.68 to 2.24; $P=0.49$) for the association with the use of cough or cold remedies that contained phenylpropanolamine, and 15.92 (95 percent confidence interval, 1.38 to 184.13; $P=0.03$) for the association with the use of appetite suppressants that contained phenylpropanolamine. An analysis in men showed no increased risk of a hemorrhagic stroke in association with the use of cough or cold remedies containing phenylpropanolamine. No men reported the use of appetite suppressants. **CONCLUSIONS:** The results suggest that phenylpropanolamine in appetite suppressants, and possibly in cough and cold remedies, is an independent risk factor for hemorrhagic stroke in women.

Question: Harm

Study Type: Case-control

Learner Level: Intermediate

Notes: This is a good example of a case Control study in the setting of a very rare outcome; The paper is a fun one because it was controversial and led to the removal of phenylpropanolamine from the market. Broadly applicable due to OTC. Good discussion points: Fun for discussing issues of applicability as there are different findings in different groups of folks (e.g. men vs. women). Might you use the same data to tell different patients different things? Other fun way to use this paper might be to consider how one would use this data to make a policy decision (e.g. in a formulary setting)

Luther, M. K., et al. (2018). "Vancomycin Plus Piperacillin-Tazobactam and Acute Kidney Injury in Adults: A Systematic Review and Meta-Analysis." *Critical Care Medicine* 46(1): 12-20.

OBJECTIVES: The objective of this systematic review and meta-analysis was to assess acute kidney injury with combination therapy of vancomycin plus piperacillin-tazobactam, in general, adult patients and in critically ill adults. Rates of acute kidney injury, time to acute kidney injury, and odds of acute kidney injury were compared with vancomycin monotherapy, vancomycin plus cefepime or carbapenem, or piperacillin-tazobactam monotherapy. **DATA SOURCES:** Studies were identified by searching Pubmed, Embase, Web of Science, and Cochrane from inception to April 2017. Abstracts from selected conference proceedings were manually searched. **STUDY SELECTION:** Articles not in English, pediatric studies, and case reports were excluded. **DATA EXTRACTION:** Two authors independently extracted data on study methods, rates of acute kidney injury, and time to acute kidney injury. Effect estimates and 95% CIs were calculated using the random effects model in RevMan 5.3. **DATA SYNTHESIS:** Literature search identified 15 published studies and 17 conference abstracts with at least 24,799 patients. The overall occurrence rate of acute kidney injury was 16.7%, with 22.2% for vancomycin plus piperacillin-tazobactam and 12.9% for comparators. This yielded an overall number needed to harm of 11. Time to acute kidney injury was faster for vancomycin plus piperacillin-tazobactam than vancomycin plus cefepime or carbapenem, but not significantly (mean difference, -1.30; 95% CI, -3.00 to 0.41 d). The odds of acute kidney injury with vancomycin plus piperacillin-tazobactam were increased versus vancomycin monotherapy (odds ratio, 3.40; 95% CI, 2.57-4.50), versus vancomycin plus cefepime or carbapenem (odds ratio, 2.68; 95% CI, 1.83-3.91), and versus piperacillin-tazobactam monotherapy (odds ratio, 2.70; 95% CI, 1.97-3.69). In a small subanalysis of 968 critically ill patients, the odds of acute kidney injury were increased versus vancomycin monotherapy (odds ratio, 9.62; 95% CI, 4.48-20.68), but not significantly different for vancomycin plus cefepime or carbapenem (odds ratio, 1.43; 95% CI, 0.83-2.47) or piperacillin-tazobactam monotherapy (odds ratio, 1.35; 95% CI, 0.86-2.11). **CONCLUSIONS:** The combination of vancomycin plus piperacillin-tazobactam increased the odds of acute kidney injury over vancomycin monotherapy, vancomycin plus cefepime or carbapenem, and piperacillin-tazobactam monotherapy. Limited data

in critically ill patients suggest the odds of acute kidney injury are increased versus vancomycin monotherapy, and mitigated versus the other comparators. Further research in the critically ill population is needed.

Question: Harm

Study Type: Meta-Analysis

Learner Level: .

Notes: .

Marcantonio, E. R., et al. (1994). "The relationship of postoperative delirium with psychoactive medications." *JAMA* 272(19): 1518-1522.

OBJECTIVE--To examine the role of medications with known psychoactive properties in the development of postoperative delirium. **DESIGN**--Nested case-control study within a prospective cohort study. **SETTING**--General surgery, orthopedic surgery, and gynecology services at Brigham and Women's Hospital, Boston, Mass. **PATIENTS**--Cases (n = 91) were patients enrolled in a prospective cohort study who developed delirium during postoperative days 2 through 5. One or two controls (n = 154) were matched to each case by the calculated preoperative risk for delirium using a predictive model developed and validated in the prospective cohort study. **MAIN OUTCOME MEASURES**--Medication exposures were ascertained from the medical record by a reviewer blinded to the study hypothesis. Exposures to narcotics, benzodiazepines, and anticholinergics were recorded for the 24-hour period before delirium developed in the 91 cases and for the same 24-hour postoperative period for the 154 matched controls. **RESULTS**--Delirium was significantly associated with postoperative exposure to meperidine (odds ratio [OR], 2.7; 95% confidence interval [CI], 1.3 to 5.5) and to benzodiazepines (OR, 3.0; 95% CI, 1.3 to 6.8). Meperidine had similar associations with delirium whether administered via epidural or patient-controlled routes, although only the epidural route reached significance (OR, 2.4; 95% CI, 1.3 to 4.4; OR, 2.1; 95% CI, 0.4 to 10.7, respectively). For benzodiazepines, long-acting agents had a trend toward stronger association with delirium than did short-acting agents (OR, 5.4; 95% CI, 1.0 to 29.2; vs 2.6; 1.1 to 6.5), and high-dose exposures had a trend toward slightly stronger association than low-dose exposures (OR, 3.3; 95% CI, 1.0 to 11.0; vs 2.6; 0.8 to 9.1). Neither narcotics (OR, 1.4; 95% CI, 0.5 to 4.3) nor anticholinergic drugs (OR, 1.5; 95% CI, 0.6 to 3.4) were significantly associated with delirium as a class, although statistical power was limited because of the high use of narcotics and the low use of anticholinergics in the study population. **CONCLUSIONS**--Clinicians caring for patients at risk for delirium should carefully evaluate the need for meperidine and benzodiazepines in the postoperative period and consider alternative therapies whenever possible.

Question: Harm

Study Type: Nested case-control

Learner Level: Beginner / Intermediate

Notes: Outstanding, prospective methodology for a case-control study (one of the best we've been able to find) Great attention to matching of cases and controls Good discussion points: Awesome paper for discussion of methods of a case control study because it is so well done and avoids some of the usual pitfalls.

Martinez, C., et al. (2005). "Antidepressant treatment and the risk of fatal and non-fatal self harm in first episode depression: nested case-control study." *BMJ* 330(7488): 389.

OBJECTIVE: To compare the risk of non-fatal self harm and suicide in patients taking selective serotonin reuptake inhibitors (SSRIs) with that of patients taking tricyclic antidepressants, as well as between different SSRIs and different tricyclic antidepressants. **DESIGN:** Nested case-control study. **SETTING:** Primary care in the United Kingdom. **PARTICIPANTS:** 146,095 individuals with a first prescription of an antidepressant for depression. **MAIN OUTCOME MEASURES:** Suicide and non-fatal self harm. **RESULTS:** 1968 cases of non-fatal self harm and 69 suicides occurred. The overall adjusted odds ratio of non-fatal self harm was 0.99 (95% confidence interval 0.86 to 1.14) and that of suicide 0.57 (0.26 to 1.25) in people prescribed SSRIs compared with those prescribed tricyclic antidepressants. We found little evidence that associations differed over time since starting or stopping treatment. We found some evidence that risks of non-fatal self harm in people prescribed SSRIs compared with those prescribed tricyclic antidepressants differed by age group (interaction P = 0.02). The adjusted odds ratio of non-fatal self harm for people prescribed SSRIs compared with users of tricyclic antidepressants for those aged 18 or younger was 1.59 (1.01 to 2.50), but no association was apparent in other age groups. No suicides occurred in those aged 18 or younger currently or recently prescribed tricyclic antidepressants or SSRIs. **CONCLUSION:** We found no evidence that the risk of suicide or non-fatal self harm in adults prescribed SSRIs was greater than in those prescribed tricyclic antidepressants. We found some weak evidence of an increased risk of non-fatal self harm for current SSRI use among those aged 18 or younger. However, preferential prescribing of SSRIs to patients at higher risk of suicidal behaviour cannot be ruled out.

Question: Harm

Study Type: Nested case-control

Learner Level: Intermediate

Notes: Well-described methodology Important topic with much controversy, widely publicized (Do SSRIs cause more suicidal behavior?); Large sample size; Paper includes clear discussion of many of the weaknesses in the methods and discussion sections; Possible Discussion Points: -Use of case-control design for rare outcome-Is this an optimal database for the study question? -Are

outcomes likely to be completely captured? ·Is there predictable bias in a specified direction? ·Did they select groups appropriately for analysis (e.g., why is venlafaxine included as "other" rather than SSRI)? ·How likely is a primary care database to include accurate diagnoses of major depressive disorder and other psychiatric disorders? ·Is it appropriate to include only those patients with depressive and related disorders while excluding those prescribed SSRIs for other or unidentified indications? ·In the multivariable model was it appropriate to include the number of covariates? (multiple comparisons) (See related article in table: Treatment of Adolescents with Depression (TADS) RCT, Therapy/Harm)

Miceli, A., et al. (2009). "Effects of angiotensin-converting enzyme inhibitor therapy on clinical outcome in patients undergoing coronary artery bypass grafting." *J Am Coll Cardiol* 54(19): 1778-1784.

OBJECTIVES: This study evaluates the effect of pre-operative angiotensin-converting enzyme inhibitor (ACEI) therapy on early clinical outcomes after coronary artery bypass grafting (CABG). **BACKGROUND:** Therapy with ACEIs has been shown to reduce the rate of mortality and prevent cardiovascular events in patients with coronary artery disease. However, their pre-operative use in patients undergoing CABG is still controversial. **METHODS:** A retrospective, observational, cohort study was undertaken of prospectively collected data on 10,023 consecutive patients undergoing isolated CABG between April 1996 and May 2008. Of these, 3,052 patients receiving pre-operative ACEI were matched to a control group by propensity score analysis. **RESULTS:** Overall rate of mortality was 1%. Pre-operative ACEI therapy was associated with a doubling in the risk of death (1.3% vs. 0.7%; odds ratio [OR]: 2.00, 95% confidence interval [CI]: 1.17 to 3.42; $p = 0.013$). There was also a significant difference between the ACEI and control group in the risk of post-operative renal dysfunction (PRD) (7.1% vs. 5.4%; OR: 1.36, 95% CI: 1.1 to 1.67; $p = 0.006$), atrial fibrillation (AF) (25% vs. 20%; OR: 1.34, 95% CI: 1.18 to 1.51; $p < 0.0001$), and increased use of inotropic support (45.9% vs. 41.1%; OR: 1.22, 95% CI: 1.1 to 1.36; $p < 0.0001$). In a multivariate analysis, pre-operative ACEI treatment was an independent predictor of mortality ($p = 0.04$), PRD ($p = 0.0002$), use of inotropic drugs ($p < 0.0001$), and AF ($p < 0.0001$). **CONCLUSIONS:** Pre-operative therapy with ACEI is associated with an increased risk of mortality, use of inotropic support, PRD, and new onset of post-operative AF.

Question: Harm

Study Type: retrospectiveCohort

Learner Level: Advanced

Notes: Good to teach propensity score and importance of randomization

Moses-Kolko, E. L., et al. (2005). "Neonatal signs after late in utero exposure to serotonin reuptake inhibitors: literature review and implications for clinical applications." *JAMA* 293(19): 2372-2383.

CONTEXT: A neonatal behavioral syndrome linked to in utero serotonin reuptake inhibitor (SRI) exposure during the last trimester of pregnancy has been identified. The US Food and Drug Administration (FDA) and drug manufacturers have recently agreed to a class labeling change for SRIs, which include selective serotonin reuptake inhibitors (SSRIs) and serotonin and norepinephrine reuptake inhibitors (SNRIs), to include information about potential adverse events in neonates exposed in utero. Integration of data about the neonatal behavioral syndrome into the management of pregnancy in women who take SRIs is a current challenge for physicians. **OBJECTIVES:** To review evidence regarding the SRI-related neonatal syndrome and to help clinicians guide their patients in a risk-benefit decision-making process. **DATA SOURCES:** We searched MEDLINE (1966-February 2005) and PsycINFO (1974-February 2005). All articles related to neonatal signs after in utero SRI exposure were acquired, as well as unpublished data on this topic from the FDA advisory committee meeting of June 2004. References cited in case reports and studies were reviewed. Foreign-language literature was included and translated to English. **STUDY SELECTION AND DATA EXTRACTION:** Studies were included if they had clearly identified maternal SRI exposure for a minimum of the final trimester of pregnancy through delivery and assessed neonatal outcomes. We identified 13 case reports describing a total of 18 cases. Nine cohort studies met criteria. When not included in the published article, relative risks and 95% confidence intervals (CIs) were computed from raw data and summary risk ratios and 95% CIs were determined with Mantel-Haenszel estimates. **DATA SYNTHESIS:** Compared with early gestational SRI exposure or no exposure, late SRI exposure carries an overall risk ratio of 3.0 (95% CI, 2.0-4.4) for a neonatal behavioral syndrome. The most SRI-related neonatal case reports involved fluoxetine and paroxetine exposures. Neonates primarily display central nervous system, motor, respiratory, and gastrointestinal signs that are usually mild and disappear by 2 weeks of age. Medical management has consisted primarily of supportive care in special care nurseries. A severe syndrome that consists of seizures, dehydration, excessive weight loss, hyperpyrexia, or intubation is rare in term infants (1/313 quantifiable cases). There have been no reported neonatal deaths attributable to neonatal SRI exposure. **CONCLUSIONS:** Available evidence indicates that in utero exposure to SRIs during the last trimester through delivery may result in a self-limited neonatal behavioral syndrome that can be managed with supportive care. The risks and benefits of discontinuing an SRI during pregnancy need to be carefully weighed for each individual patient. Development and validation of assessment methods and clinical management strategies are critical to advancing this research.

Question: Harm

Study Type: Systematic Review

Learner Level: Moderate ot Advanced

Notes: Good article if you want to expose learners to systematic review addressing harm instead of therapy

Murphy, T. V., et al. (2001). "Intussusception among infants given an oral rotavirus vaccine." *N Engl J Med* 344(8): 564-572.

BACKGROUND: Intussusception is a form of intestinal obstruction in which a segment of the bowel prolapses into a more distal segment. Our investigation began on May 27, 1999, after nine cases of infants who had intussusception after receiving the tetraivalent rhesus-human reassortant rotavirus vaccine (RRV-TV) were reported to the Vaccine Adverse Event Reporting System. **METHODS:** In 19 states, we assessed the potential association between RRV-TV and intussusception among infants at least 1 but less than 12 months old. Infants hospitalized between November 1, 1998, and June 30, 1999, were identified by systematic reviews of medical and radiologic records. Each infant with intussusception was matched according to age with four healthy control infants who had been born at the same hospital as the infant with intussusception. Information on vaccinations was verified by the provider. **RESULTS:** Data were analyzed for 429 infants with intussusception and 1763 matched controls in a case-control analysis as well as for 432 infants with intussusception in a case-series analysis. Seventy-four of the 429 infants with intussusception (17.2 percent) and 226 of the 1763 controls (12.8 percent) had received RRV-TV ($P=0.02$). An increased risk of intussusception 3 to 14 days after the first dose of RRV-TV was found in the case-control analysis (adjusted odds ratio, 21.7; 95 percent confidence interval, 9.6 to 48.9). In the case-series analysis, the incidence-rate ratio was 29.4 (95 percent confidence interval, 16.1 to 53.6) for days 3 through 14 after a first dose. There was also an increase in the risk of intussusception after the second dose of the vaccine, but it was smaller than the increase in risk after the first dose. Assuming full implementation of a national program of vaccination with RRV-TV, we estimated that 1 case of intussusception attributable to the vaccine would occur for every 4670 to 9474 infants vaccinated. **CONCLUSIONS:** The strong association between vaccination with RRV-TV and intussusception among otherwise healthy infants supports the existence of a causal relation. Rotavirus vaccines with an improved safety profile are urgently needed.

Question: Harm

Study Type: Case-control

Learner Level: Intermediate

Notes: A well done case-controlled study in pediatrics with clear methodology and results. Good discussion points: Provides a basis for discussion of case-controlled methods and post-marketing surveillance data for investigating harm from preventative (i.e. vaccine) or therapeutic interventions.

Newman, L. S., et al. (2004). "A case control etiologic study of sarcoidosis: environmental and occupational risk factors." *Am J Respir Crit Care Med* 170(12): 1324-1330.

Past research suggests that environmental factors may be associated with sarcoidosis risk. We conducted a case control study to test a priori hypotheses that environmental and occupational exposures are associated with sarcoidosis. Ten centers recruited 706 newly diagnosed patients with sarcoidosis and an equal number of age-, race-, and sex-matched control subjects. Interviewers administered questionnaires containing questions regarding occupational and nonoccupational exposures that we assessed in univariable and multivariable analyses. We observed positive associations between sarcoidosis and specific occupations (e.g., agricultural employment, odds ratio [OR] 1.46, confidence interval [CI] 1.13-1.89), exposures (e.g., insecticides at work, OR 1.52, CI 1.14-2.04, and work environments with mold/mildew exposures [environments with possible exposures to microbial bioaerosols], OR 1.61, CI 1.13-2.31). A history of ever smoking cigarettes was less frequent among cases than control subjects (OR 0.62, CI 0.50-0.77). In multivariable modeling, we observed elevated ORs for work in areas with musty odors (OR 1.62, CI 1.24-2.11) and with occupational exposure to insecticides (OR 1.61, CI 1.13-2.28), and a decreased OR related to ever smoking cigarettes (OR 0.65, CI 0.51-0.82). The study did not identify a single, predominant cause of sarcoidosis. We identified several exposures associated with sarcoidosis risk, including insecticides, agricultural employment, and microbial bioaerosols.

Question: Harm

Study Type: Case-Control Study

Learner Level: Intermediate

Notes: Good for helping learners understand how to read a case-control study, how to think about adjusted analysis and odds ratios; paper does not lend itself to calculating odds ratios (which is okay!).

Rugbjerg, K., et al. (2008). "Risk of Parkinson's disease after hospital contact for head injury: population based case-control study." *BMJ* 337: a2494.

OBJECTIVE: To determine whether a hospital contact for a head injury increases the risk of subsequently developing Parkinson's disease. **DESIGN:** Population based case-control study. **SETTING:** Denmark. **PARTICIPANTS:** 13 695 patients with a primary diagnosis of Parkinson's disease in the Danish national hospital register during 1986-2006, each matched on age and sex to five population controls selected at random from inhabitants in Denmark alive at the date of the patient's diagnosis (n=68 445). **MAIN OUTCOME MEASURES:** Hospital contacts for head injuries ascertained from hospital register; frequency of history of head injury. **RESULTS:** An overall 50% increase in prevalence of hospital contacts for head injury was seen before the first registration of Parkinson's disease in this population (odds ratio 1.5, 95% confidence interval 1.4 to 1.7). The observed association was, however, due almost entirely to injuries that occurred during the three months before the first record of Parkinson's disease (odds ratio 8.0, 5.6 to 11.6), and no association was found between the two events when they occurred 10 or more years apart (1.1, 0.9 to 1.3). **CONCLUSIONS:** The steeply increased frequency of hospital contacts for a head injury during the months preceding the date at which Parkinson's disease was first recorded is a consequence of the evolving movement disorder rather than its cause.

Question: Harm

Study Type: Case-Control

Learner Level: Beginner

Notes: It is easy to catch students' attention at the introduction showing pictures of Muhammad Ali, and b) it illustrates very well the concept of reverse causality.

Schilcher, J., et al. (2011). "Bisphosphonate use and atypical fractures of the femoral shaft." *N Engl J Med* 364(18): 1728-1737.

BACKGROUND: Studies show conflicting results regarding the possible excess risk of atypical fractures of the femoral shaft associated with bisphosphonate use. **METHODS:** In Sweden, 12,777 women 55 years of age or older sustained a fracture of the femur in 2008. We reviewed radiographs of 1234 of the 1271 women who had a subtrochanteric or shaft fracture and identified 59 patients with atypical fractures. Data on medications and coexisting conditions were obtained from national registries. The relative and absolute risk of atypical fractures associated with bisphosphonate use was estimated by means of a nationwide cohort analysis. The 59 case patients were also compared with 263 control patients who had ordinary subtrochanteric or shaft fractures. **RESULTS:** The age-adjusted relative risk of atypical fracture was 47.3 (95% confidence interval [CI], 25.6 to 87.3) in the cohort analysis. The increase in absolute risk was 5 cases per 10,000 patient-years (95% CI, 4 to 7). A total of 78% of the case patients and 10% of the controls had received bisphosphonates, corresponding to a multivariable-adjusted odds ratio of 33.3 (95% CI, 14.3 to 77.8). The risk was independent of coexisting conditions and of concurrent use of other drugs with known effects on bone. The duration of use influenced the risk (odds ratio per 100 daily doses, 1.3; 95% CI, 1.1 to 1.6). After drug withdrawal, the risk diminished by 70% per year since the last use (odds ratio, 0.28; 95% CI, 0.21 to 0.38). **CONCLUSIONS:** These population-based nationwide analyses may be reassuring for patients who receive bisphosphonates. Although there was a high prevalence of current bisphosphonate use among patients with atypical fractures, the absolute risk was small. (Funded by the Swedish Research Council.).

Question: Harm

Study Type: Cohort Study

Learner Level: Beginner

Notes: Harm – well done cohort study detecting rare, duration-dependent event (atypical femoral fractures with bisphosphonates for osteoporosis treatment) great to compare/contrast a cohort with RR reported and case control with OR reported in the same paper, and great to point out that even strong associations may not change practice if events are very rare. great for creating 2x2 tables of both the cohort numbers and the case control numbers to compare and contrast; great example of huge relative association having little bearing on absolute risks, since risk is tiny. Easy to find the relevant numbers for calculating RR and OR.

Schjerning Olsen, A. M., et al. (2011). "Duration of treatment with nonsteroidal anti-inflammatory drugs and impact on risk of death and recurrent myocardial infarction in patients with prior myocardial infarction: a nationwide cohort study." *Circulation* 123(20): 2226-2235.

BACKGROUND: Despite the fact that nonsteroidal anti-inflammatory drugs (NSAIDs) are contraindicated among patients with established cardiovascular disease, many receive NSAID treatment for a short period of time. However, little is known about the association between NSAID treatment duration and risk of cardiovascular disease. We therefore studied the duration of NSAID treatment and cardiovascular risk in a nationwide cohort of patients with prior myocardial infarction (MI). **METHODS AND RESULTS:** Patients ≥ 30 years of age who were admitted with first-time MI during 1997 to 2006 and their subsequent NSAID use were identified by individual-level linkage of nationwide registries of hospitalization and drug dispensing from pharmacies in Denmark. Risk of death and recurrent MI according to duration of NSAID treatment was analyzed by multivariable time-stratified Cox proportional-hazard models and by incidence rates per 1000 person-years. Of the 83 677 patients included, 42.3% received NSAIDs during follow-up. There were 35 257 deaths/recurrent MIs. Overall, NSAID treatment was significantly associated with an increased risk of death/recurrent MI (hazard ratio, 1.45; 95% confidence interval, 1.29 to 1.62) at the beginning of the treatment, and the risk persisted throughout the treatment course (hazard ratio, 1.55; 95% confidence interval, 1.46 to 1.64 after 90 days). Analyses of individual NSAIDs showed that the traditional NSAID diclofenac was associated with the highest risk (hazard ratio, 3.26; 95% confidence interval, 2.57 to 3.86 for death/MI at day 1 to 7 of treatment). **CONCLUSIONS:** Even short-term treatment with most NSAIDs was associated with increased risk of death and recurrent MI in patients with prior MI. Neither short- nor long-term treatment with NSAIDs is advised in this population, and any NSAID use should be limited from a cardiovascular safety point of view.

Question: Harm

Study Type: Cohort Study

Learner Level: Intermediate

Notes: cohort to review HARM - Denmark, great cohort, but some issues with validity/sample sizes/confounding by indication, as well as small absolute risk of harm.

Silverstein, F. E., et al. (2000). "Gastrointestinal toxicity with celecoxib vs nonsteroidal anti-inflammatory drugs for osteoarthritis and rheumatoid arthritis: the CLASS study: A randomized controlled trial. Celecoxib Long-term Arthritis Safety Study." *JAMA* 284(10): 1247-1255.

CONTEXT: Conventional nonsteroidal anti-inflammatory drugs (NSAIDs) are associated with a spectrum of toxic effects, notably gastrointestinal (GI) effects, because of inhibition of cyclooxygenase (COX)-1. Whether COX-2-specific inhibitors are associated with fewer clinical GI toxic effects is unknown. **OBJECTIVE:** To determine whether celecoxib, a COX-2-specific inhibitor, is associated with

a lower incidence of significant upper GI toxic effects and other adverse effects compared with conventional NSAIDs. DESIGN: The Celecoxib Long-term Arthritis Safety Study (CLASS), a double-blind, randomized controlled trial conducted from September 1998 to March 2000. SETTING: Three hundred eighty-six clinical sites in the United States and Canada. PARTICIPANTS: A total of 8059 patients (>=18 years old) with osteoarthritis (OA) or rheumatoid arthritis (RA) were enrolled in the study, and 7968 received at least 1 dose of study drug. A total of 4573 patients (57%) received treatment for 6 months. INTERVENTIONS: Patients were randomly assigned to receive celecoxib, 400 mg twice per day (2 and 4 times the maximum RA and OA dosages, respectively; n = 3987); ibuprofen, 800 mg 3 times per day (n = 1985); or diclofenac, 75 mg twice per day (n = 1996). Aspirin use for cardiovascular prophylaxis (</=325 mg/d) was permitted. MAIN OUTCOME MEASURES: Incidence of prospectively defined symptomatic upper GI ulcers and ulcer complications (bleeding, perforation, and obstruction) and other adverse effects during the 6-month treatment period. RESULTS: For all patients, the annualized incidence rates of upper GI ulcer complications alone and combined with symptomatic ulcers for celecoxib vs NSAIDs were 0.76% vs 1.45% (P =.09) and 2.08% vs 3.54% (P =.02), respectively. For patients not taking aspirin, the annualized incidence rates of upper GI ulcer complications alone and combined with symptomatic ulcers for celecoxib vs NSAIDs were 0.44% vs 1.27% (P =.04) and 1.40% vs 2.91% (P =.02). For patients taking aspirin, the annualized incidence rates of upper GI ulcer complications alone and combined with symptomatic ulcers for celecoxib vs NSAIDs were 2.01% vs 2.12% (P =.92) and 4.70% vs 6.00% (P =.49). Fewer celecoxib-treated patients than NSAID-treated patients experienced chronic GI blood loss, GI intolerance, hepatotoxicity, or renal toxicity. No difference was noted in the incidence of cardiovascular events between celecoxib and NSAIDs, irrespective of aspirin use. CONCLUSIONS: In this study, celecoxib, at dosages greater than those indicated clinically, was associated with a lower incidence of symptomatic ulcers and ulcer complications combined, as well as other clinically important toxic effects, compared with NSAIDs at standard dosages. The decrease in upper GI toxicity was strongest among patients not taking aspirin concomitantly. JAMA. 2000;284:1247-1255

Question: Harm

Study Type: RCT

Learner Level: Intermediate / Advanced

Notes: This is a sample packet. See the entire teaching package that is in this section. This paper is an example of evaluation of Harm using RCT methodology. It is an opportunity to discuss that there are multiple methodologies that can be used for each kind of question. Beware, however, that questions of harm are frequently not answered by RCT (if the outcomes are too rare or if there are ethical issues of randomization when known harms are involved.) Good discussion points: See sample package for using this paper to bring up some issues of ethics in medical reporting and drug company sponsorship of clinical trials.

Wadhwa, R. K., et al. (2018). "Association of the Hospital Readmissions Reduction Program With Mortality Among Medicare Beneficiaries Hospitalized for Heart Failure, Acute Myocardial Infarction, and Pneumonia." *JAMA* 320(24): 2542-2552.

Importance: The Hospital Readmissions Reduction Program (HRRP) has been associated with a reduction in readmission rates for heart failure (HF), acute myocardial infarction (AMI), and pneumonia. It is unclear whether the HRRP has been associated with change in patient mortality. Objective: To determine whether the HRRP was associated with a change in patient mortality. Design, Setting, and Participants: Retrospective cohort study of hospitalizations for HF, AMI, and pneumonia among Medicare fee-for-service beneficiaries aged at least 65 years across 4 periods from April 1, 2005, to March 31, 2015. Period 1 and period 2 occurred before the HRRP to establish baseline trends (April 2005-September 2007 and October 2007-March 2010). Period 3 and period 4 were after HRRP announcement (April 2010 to September 2012) and HRRP implementation (October 2012 to March 2015). Exposures: Announcement and implementation of the HRRP. Main Outcomes and Measures: Inverse probability-weighted mortality within 30 days of discharge following hospitalization for HF, AMI, and pneumonia, and stratified by whether there was an associated readmission. An additional end point was mortality within 45 days of initial hospital admission for target conditions. Results: The study cohort included 8.3 million hospitalizations for HF, AMI, and pneumonia, among which 7.9 million (mean age, 79.6 [8.7] years; 53.4% women) were alive at discharge. There were 3.2 million hospitalizations for HF, 1.8 million for AMI, and 3.0 million for pneumonia. There were 270517 deaths within 30 days of discharge for HF, 128088 for AMI, and 246154 for pneumonia. Among patients with HF, 30-day postdischarge mortality increased before the announcement of the HRRP (0.27% increase from period 1 to period 2). Compared with this baseline trend, HRRP announcement (0.49% increase from period 2 to period 3; difference in change, 0.22%, P = .01) and implementation (0.52% increase from period 3 to period 4; difference in change, 0.25%, P = .001) were significantly associated with an increase in postdischarge mortality. Among patients with AMI, HRRP announcement was associated with a decline in postdischarge mortality (0.18% pre-HRRP increase vs 0.08% post-HRRP announcement decrease; difference in change, -0.26%; P = .01) and did not significantly change after HRRP implementation. Among patients with pneumonia, postdischarge mortality was stable before HRRP (0.04% increase from period 1 to period 2), but significantly increased after HRRP announcement (0.26% post-HRRP announcement increase; difference in change, 0.22%, P = .01) and implementation (0.44% post-HRRP implementation increase; difference in change, 0.40%, P < .001). The overall increase in mortality among patients with HF and pneumonia was mainly related to outcomes among patients who were not readmitted but died within 30 days of discharge. For all 3 conditions, HRRP implementation was not significantly associated with an increase in mortality within 45 days of admission, relative to pre-HRRP trends. Conclusions and Relevance: Among Medicare beneficiaries, the HRRP was significantly associated with an increase in 30-day postdischarge mortality after hospitalization for HF and pneumonia, but not for AMI. Given the study design and the lack of significant association of the HRRP with mortality within 45 days of admission, further research is needed to understand whether the increase in 30-day postdischarge mortality is a result of the policy.

Question: Harm

Study Type: Retrospective Cohort Study

Learner Level: intermediate

Notes: A really interesting paper looking at a systems intervention (hospital readmissions reduction program) that did decrease readmission (an intermediate outcome) but increased mortality (a patient important outcome). Great for discussion of the difference between association and causality. It is a well done retrospective cohort with a very large sample size. There is a useful editorial (by Greff Fonarow) in the same issue of JAMA.

Wang, P. S., et al. (2005). "Risk of death in elderly users of conventional vs. atypical antipsychotic medications." *N Engl J Med* 353(22): 2335-2341.

BACKGROUND: Recently, the Food and Drug Administration (FDA) issued an advisory stating that atypical antipsychotic medications increase mortality among elderly patients. However, the advisory did not apply to conventional antipsychotic medications; the risk of death with these older agents is not known. **METHODS:** We conducted a retrospective cohort study involving 22,890 patients 65 years of age or older who had drug insurance benefits in Pennsylvania and who began receiving a conventional or atypical antipsychotic medication between 1994 and 2003. Analyses of mortality rates and Cox proportional-hazards models were used to compare the risk of death within 180 days, less than 40 days, 40 to 79 days, and 80 to 180 days after the initiation of therapy with an antipsychotic medication. We controlled for potential confounding variables with the use of traditional multivariate Cox models, propensity-score adjustments, and an instrumental-variable analysis. **RESULTS:** Conventional antipsychotic medications were associated with a significantly higher adjusted risk of death than were atypical antipsychotic medications at all intervals studied (< or =180 days: relative risk, 1.37; 95 percent confidence interval, 1.27 to 1.49; <40 days: relative risk, 1.56; 95 percent confidence interval, 1.37 to 1.78; 40 to 79 days: relative risk, 1.37; 95 percent confidence interval, 1.19 to 1.59; and 80 to 180 days: relative risk, 1.27; 95 percent confidence interval, 1.14 to 1.41) and in all subgroups defined according to the presence or absence of dementia or nursing home residency. The greatest increases in risk occurred soon after therapy was initiated and with higher dosages of conventional antipsychotic medications. Increased risks associated with conventional as compared with atypical antipsychotic medications persisted in confirmatory analyses performed with the use of propensity-score adjustment and instrumental-variable estimation. **CONCLUSIONS:** If confirmed, these results suggest that conventional antipsychotic medications are at least as likely as atypical agents to increase the risk of death among elderly persons and that conventional drugs should not be used to replace atypical agents discontinued in response to the FDA warning.

Question: Harm

Study Type: Cohort, retrospective

Learner Level: Intermediate / Advanced

Notes: Important clinical question relevant to generalists and geriatricians regarding commonplace treatment practices (administration of antipsychotics to demented elderly patients) now thought to cause harm. Administrative databases capture most of the elderly population and include complete mortality data based on medicare and social security. Clearly defined exposure groups and methodology Appropriate modeling after initial analysis includes utilization of propensity scores, instrumental-variable analysis and modeling strategies to control for possible biases in what doctors prescribed in the first place, with sensitivity analysis also conducted. Good discussion points: How does one confirm cause and effect in a retrospective cohort study? Validity criteria for "harm" articles (more advanced): use of propensity score and adjustment for variables that may have contributed to bias in prescribing patterns. Given the information in this article in context of limited information for effectiveness of any intervention for agitation in dementia...what would participants choose to do in order to treat behavioral symptoms in dementia?

Williams, C., et al. (1998). "Case-control study of salmeterol and near-fatal attacks of asthma." *Thorax* 53(1): 7-13.

BACKGROUND: A case-control study was undertaken to investigate the hypothesis that the use of the long acting beta agonist salmeterol increases the risk of a near-fatal attack of asthma. **METHODS:** The cases comprised admissions to the intensive care unit (ICU) for asthma in 14 major hospitals within the Wessex region in 1992. For each of the cases four age-matched controls were selected from asthma admissions to the same hospital during the same period. Information on prescribed drug therapy for the 48 cases and 185 controls was collected from the hospital admission records. **RESULTS:** The patients admitted to the ICU had greater chronic asthma severity and had generally been prescribed more asthma drugs than the control admissions to hospital. The relative risk of a near-fatal attack of asthma in patients prescribed inhaled salmeterol was 2.32 (95% CI 1.05 to 5.16), $p = 0.04$. However, the salmeterol relative risk decreased to 1.42 (95% CI 0.49 to 4.10), $p = 0.52$ when the analysis was restricted to the more chronically severe patients (those in the subgroup of patients with a hospital admission for asthma in the previous 12 months). These findings suggest that the increased unadjusted relative risk with salmeterol is predominantly due to confounding by severity—that is, the increased relative risk is due to patients with more severe asthma (at greatest risk of a near-fatal asthma attack) being preferentially prescribed salmeterol. This interpretation is supported by the finding in this study that, within the control group (selected from the population of asthmatics requiring hospital admission), salmeterol was preferentially prescribed to the most severe patients (a threefold greater prescription of salmeterol to control patients if they had been admitted to hospital in the 12 months prior to the index admission). There was no increased risk of a near-fatal attack of asthma in patients prescribed a beta agonist by metered dose inhaler (OR 0.75 (95% CI 0.31 to 1.78), $p = 0.51$). In contrast, the relative risks for beta agonists delivered by nebulisation (OR 3.86 (95% CI 1.99 to 7.50), $p < 0.001$) and oral theophylline (OR 2.45 (95% CI 1.26 to 4.78), $p < 0.01$) were increased and did not markedly decrease when the analysis was restricted to the more severe asthmatic subjects. **CONCLUSIONS:** Although these findings are not conclusive, particularly because of the small numbers involved in some subgroup analyses, they suggest that the use of salmeterol by patients with chronic severe asthma is not associated with a significantly increased risk of a near-fatal attack of asthma. If a near-fatal asthma attack is considered to be an intermediate step in a process by which a severe attack of asthma may become fatal, these results would suggest that salmeterol is unlikely to be associated with an increased risk of death, at least by this mechanism.

Question: Harm

Study Type: Case-control

Learner Level: Advanced

Notes: The FDA Talk Paper August 14, 2003 and the Health and Drug Alert that was issued by the Canadian Medical Association in JAMC both allude to data from the SMART trial (Salmeterol Multi-Center Asthma Research Trial) that was stopped in interim analysis in 2003. Interim analysis in 2002 showed a non-significant trend toward more asthma related life threatening events in all patients and a significant increase in the subgroup of black patients. The SMART trial prompted the drug warning for salmeterol. However, the SMART trial remains unpublished. Thus, the Williams study (case-control) is the best evidence actually published. Good discussion points: Used in conjunction with FDA Talk Paper Aug 14, 2003 regarding the SMART trial or other (GSK warning insert, Micromedex adverse reactions, etc.) makes for an interesting discussion regarding harm and finding the evidence- Where is the publication of SMART (Salmeterol Multi-Center Asthma Research Trial)? Why might the drug company sponsors not have published it? Other discussion points: Surrogate endpoints Use of both 'odds ratio' and 'risk ratio' language. Why does this matter for a case-control study? Is extracting drug data from the chart an accurate method of determining who was and was not actually taking salmeterol? Why might the Case-Control data possibly differ from the SMART data (at least the data we do know)? See also: Wooltorton 2003 and FDA Talk Paper - Labeling changes for drug products that contain salmeterol (<http://www.fda.gov/bbs/topics/ANSWERS/2003/ANS01248.html> viewed 12/28/05)

Wooltorton, E. (2003). "Salmeterol (Serevent) asthma trial halted early." CMAJ 168(6): 738.

Question: Harm

Study Type: Case-control

Learner Level: Advanced

Notes: The FDA Talk Paper August 14, 2003 and the Health and Drug Alert that was issued by the Canadian Medical Association in JAMC both allude to data from the SMART trial (Salmeterol Multi-Center Asthma Research Trial) that was stopped in interim analysis in 2003. Interim analysis in 2002 showed a non-significant trend toward more asthma related life threatening events in all patients and a significant increase in the subgroup of black patients. The SMART trial prompted the drug warning for salmeterol. However, the SMART trial remains unpublished. Thus, the Williams study (case-control) is the best evidence actually published. Good discussion points: Used in conjunction with FDA Talk Paper Aug 14, 2003 regarding the SMART trial or other (GSK warning insert, Micromedex adverse reactions, etc.) makes for an interesting discussion regarding harm and finding the evidence- Where is the publication of SMART (Salmeterol Multi-Center Asthma Research Trial)? Why might the drug company sponsors not have published it? Other discussion points: Surrogate endpoints Use of both 'odds ratio' and 'risk ratio' language. Why does this matter for a case-control study? Is extracting drug data from the chart an accurate method of determining who was and was not actually taking salmeterol? Why might the Case-Control data possibly differ from the SMART data (at least the data we do know)? See also: Williams 2003 and FDA Talk Paper - Labeling changes for drug products that contain salmeterol (<http://www.fda.gov/bbs/topics/ANSWERS/2003/ANS01248.html> viewed 12/28/05)

Wysowski, D. K., et al. (2001). "An analysis of reports of depression and suicide in patients treated with isotretinoin." Journal of the American Academy of Dermatology 45(4): 515-519.

BACKGROUND: The Food and Drug Administration (FDA) has received reports of depression and suicide in patients treated with isotretinoin. **OBJECTIVE:** Our purpose was to provide the number and describe the cases of depression and suicide reported to the FDA in US patients treated with isotretinoin and to consider the nature of a possible association between isotretinoin and depression. **METHODS:** An analysis was made of reports of depression, suicidal ideation, suicide attempt, and suicide in US isotretinoin users voluntarily submitted to the manufacturer and the FDA from 1982 to May 2000 and entered in the FDA's Adverse Event Reporting System database. **RESULTS:** From marketing of isotretinoin in 1982 to May 2000, the FDA received reports of 37 US patients treated with isotretinoin who committed suicide; 110 who were hospitalized for depression, suicidal ideation, or suicide attempt; and 284 with nonhospitalized depression, for a total of 431 patients. Factors suggesting a possible association between isotretinoin and depression include a temporal association between use of the drug and depression, positive dechallenges (often with psychiatric treatment), positive rechallenges, and possible biologic plausibility. Compared with all drugs in the FDA's Adverse Event Reporting System database to June 2000, isotretinoin ranked within the top 10 for number of reports of depression and suicide attempt. **CONCLUSION:** The FDA has received reports of depression, suicidal ideation, suicide attempt, and suicide in patients treated with isotretinoin. Additional studies are needed to determine whether isotretinoin causes depression and to identify susceptible persons. In the meantime, physicians are advised to inform patients prescribed isotretinoin (and parents, if appropriate) of the possibility of development or worsening of depression. They should advise patients (and parents) to immediately report mood swings and symptoms suggestive of depression such as sadness, crying, loss of appetite, unusual fatigue, withdrawal, and inability to concentrate so that patients can be promptly evaluated for appropriate treatment, including consideration of drug discontinuation and referral for psychiatric care.

Question: Harm

Study Type: Case Series of FDA reports

Learner Level: Beginner / Intermediate

Notes: This paper is specifically good for discussion regarding how to make decisions about potential harms of therapy when there are no controlled trials available. Good discussion points: Can use this as a jumping point to discuss the 'hierarchy of evidence' and to drive home that, in the end, you need to make clinical decisions with whatever evidence is available.

Yang, Y. X., et al. (2006). "Long-term proton pump inhibitor therapy and risk of hip fracture." *JAMA* 296(24): 2947-2953.

CONTEXT: Proton pump inhibitors (PPIs) may interfere with calcium absorption through induction of hypochlorhydria but they also may reduce bone resorption through inhibition of osteoclastic vacuolar proton pumps. **OBJECTIVE:** To determine the association between PPI therapy and risk of hip fracture. **DESIGN, SETTING, AND PATIENTS:** A nested case-control study was conducted using the General Practice Research Database (1987-2003), which contains information on patients in the United Kingdom. The study cohort consisted of users of PPI therapy and nonusers of acid suppression drugs who were older than 50 years. Cases included all patients with an incident hip fracture. Controls were selected using incidence density sampling, matched for sex, index date, year of birth, and both calendar period and duration of up-to-standard follow-up before the index date. For comparison purposes, a similar nested case-control analysis for histamine 2 receptor antagonists was performed. **MAIN OUTCOME MEASURE:** The risk of hip fractures associated with PPI use. **RESULTS:** There were 13,556 hip fracture cases and 135,386 controls. The adjusted odds ratio (AOR) for hip fracture associated with more than 1 year of PPI therapy was 1.44 (95% confidence interval [CI], 1.30-1.59). The risk of hip fracture was significantly increased among patients prescribed long-term high-dose PPIs (AOR, 2.65; 95% CI, 1.80-3.90; $P < .001$). The strength of the association increased with increasing duration of PPI therapy (AOR for 1 year, 1.22 [95% CI, 1.15-1.30]; 2 years, 1.41 [95% CI, 1.28-1.56]; 3 years, 1.54 [95% CI, 1.37-1.73]; and 4 years, 1.59 [95% CI, 1.39-1.80]; $P < .001$ for all comparisons). **CONCLUSION:** Long-term PPI therapy, particularly at high doses, is associated with an increased risk of hip fracture.

Study Type: Nested Case Control

Learner Level: Advanced

Notes: Nested case control study looking at association between PPI treatment and hip fracture. Good methodology for explaining a nested case control design.

Other

Costantino, G., et al. (2018). "Prevalence of Pulmonary Embolism in Patients With Syncope." *JAMA Intern Med* 178(3): 356-362.

Importance: Sparse data and conflicting evidence exist on the prevalence of pulmonary embolism (PE) in patients with syncope. **Objective:** To estimate the prevalence of PE among patients presenting to the emergency department (ED) for evaluation of syncope. **Design, Setting, and Participants:** This retrospective, observational study analyzed longitudinal administrative data from 5 databases in 4 different countries (Canada, Denmark, Italy, and the United States). Data from all adult patients (aged ≥ 18 years) who presented to the ED were screened to identify those with syncope codes at discharge. Data were collected from January 1, 2000, through September 30, 2016. **Main Outcomes and Measures:** The prevalence of PE at ED and hospital discharge, identified using codes from the International Classification of Diseases, was considered the primary outcome. Two sensitivity analyses considering prevalence of PE at 90 days of follow-up and prevalence of venous thromboembolism were performed. **Results:** A total of 1671944 unselected adults who presented to the ED for syncope were included. The prevalence of PE, according to administrative data, ranged from 0.06% (95% CI, 0.05%-0.06%) to 0.55% (95% CI, 0.50%-0.61%) for all patients and from 0.15% (95% CI, 0.14%-0.16%) to 2.10% (95% CI, 1.84%-2.39%) for hospitalized patients. The prevalence of PE at 90 days of follow-up ranged from 0.14% (95% CI, 0.13%-0.14%) to 0.83% (95% CI, 0.80%-0.86%) for all patients and from 0.35% (95% CI, 0.34%-0.37%) to 2.63% (95% CI, 2.34%-2.95%) for hospitalized patients. Finally, the prevalence of venous thromboembolism at 90 days ranged from 0.30% (95% CI, 0.29%-0.31%) to 1.37% (95% CI, 1.33%-1.41%) for all patients and from 0.75% (95% CI, 0.73%-0.78%) to 3.86% (95% CI, 3.51%-4.24%) for hospitalized patients. **Conclusions and Relevance:** Pulmonary embolism was rarely identified in patients with syncope. Although PE should be considered in every patient, not all patients should undergo evaluation for PE.

Question: Disease probability

Study Type: Retrospective Cohort Study

Learner Level: Intermediate

Notes: Good in combination with article by Prandoni on the same topic to compare and contrast methods of determining disease probability. The Costantino article demonstrates how data is obtained from administrative databases whereas the Prandoni paper is a prospective study with clinically recruited patients.

Cryer, P. E., et al. (2009). "Evaluation and management of adult hypoglycemic disorders: an Endocrine Society Clinical Practice Guideline." *J Clin Endocrinol Metab* 94(3): 709-728.

OBJECTIVE: The aim is to provide guidelines for the evaluation and management of adults with hypoglycemic disorders, including those with diabetes mellitus. **EVIDENCE:** Using the recommendations of the Grading of Recommendations, Assessment, Development and Evaluation (GRADE) system, the quality of evidence is graded very low (plus sign in circle ooo), low (plus sign in circle plus sign in circle oo), moderate (plus sign in circle plus sign in circle plus sign in circle o), or high (plus sign in circle plus sign in circle plus sign in circle plus sign in circle). **CONCLUSIONS:** We recommend evaluation and management of hypoglycemia only in patients in whom Whipple's triad--symptoms, signs, or both consistent with hypoglycemia, a low plasma glucose concentration, and resolution of those symptoms or signs after the plasma glucose concentration is raised--is documented. In patients with hypoglycemia without diabetes mellitus, we recommend the following strategy. First, pursue clinical clues to potential hypoglycemic

etiologies--drugs, critical illnesses, hormone deficiencies, nonislet cell tumors. In the absence of these causes, the differential diagnosis narrows to accidental, surreptitious, or even malicious hypoglycemia or endogenous hyperinsulinism. In patients suspected of having endogenous hyperinsulinism, measure plasma glucose, insulin, C-peptide, proinsulin, beta-hydroxybutyrate, and circulating oral hypoglycemic agents during an episode of hypoglycemia and measure insulin antibodies. Insulin or insulin secretagogue treatment of diabetes mellitus is the most common cause of hypoglycemia. We recommend the practice of hypoglycemia risk factor reduction--addressing the issue of hypoglycemia, applying the principles of intensive glycemic therapy, and considering both the conventional risk factors and those indicative of compromised defenses against falling plasma glucose concentrations--in persons with diabetes.

Question: Therapy

Study Type: Guideline

Learner Level: Intermediate

Notes: Hypoglycemia guideline: GRADE is emerging as the most coherent and comprehensive approach to guideline development, healthcare providers need to become fluent in guideline appraisal and interpretation using GRADE-developed CPGs.

Hagan, P. G., et al. (2000). "The International Registry of Acute Aortic Dissection (IRAD): new insights into an old disease." *JAMA* 283(7): 897-903.

CONTEXT: Acute aortic dissection is a life-threatening medical emergency associated with high rates of morbidity and mortality. Data are limited regarding the effect of recent imaging and therapeutic advances on patient care and outcomes in this setting. **OBJECTIVE:** To assess the presentation, management, and outcomes of acute aortic dissection. **DESIGN:** Case series with patients enrolled between January 1996 and December 1998. Data were collected at presentation and by physician review of hospital records. **SETTING:** The International Registry of Acute Aortic Dissection, consisting of 12 international referral centers. **PARTICIPANTS:** A total of 464 patients (mean age, 63 years; 65.3% male), 62.3% of whom had type A dissection. **MAIN OUTCOME MEASURES:** Presenting history, physical findings, management, and mortality, as assessed by history and physician review of hospital records. **RESULTS:** While sudden onset of severe sharp pain was the single most common presenting complaint, the clinical presentation was diverse. Classic physical findings such as aortic regurgitation and pulse deficit were noted in only 31.6% and 15.1% of patients, respectively, and initial chest radiograph and electrocardiogram were frequently not helpful (no abnormalities were noted in 12.4% and 31.3% of patients, respectively). Computed tomography was the initial imaging modality used in 61.1%. Overall in-hospital mortality was 27.4%. Mortality of patients with type A dissection managed surgically was 26%; among those not receiving surgery (typically because of advanced age and comorbidity), mortality was 58%. Mortality of patients with type B dissection treated medically was 10.7%. Surgery was performed in 20% of patients with type B dissection; mortality in this group was 31.4%. **CONCLUSIONS:** Acute aortic dissection presents with a wide range of manifestations, and classic findings are often absent. A high clinical index of suspicion is necessary. Despite recent advances, in-hospital mortality rates remain high. Our data support the need for continued improvement in prevention, diagnosis, and management of acute aortic dissection.

Question: Clinical manifestations of disease

Study Type: Case Series

Learner Level: Beginner

Notes: A study of the frequency of clinical manifestations of disease, specifically of aortic dissection – this registry study of a less common yet dramatic disorder also gets clinical learner's attention, and illustrates the need for thorough and consistent diagnostic evaluation and the challenges of applicability across types of clinical settings and across time.

Hart, R. (2006). "A simple risk score predicted 7-day stroke risk after transient ischemic attack." *ACP Journal Club* 144(1): 24.

Question: Clinical Prediction Guideline

Study Type: Observational

Learner Level: Intermediate

Notes: Highly applicable to internal medicine; No illustration of variability in results of cohorts

Kelly, A. M., et al. (2017). "An Observational Study of Dyspnea in Emergency Departments: The Asia, Australia, and New Zealand Dyspnea in Emergency Departments Study (AANZDEM)." *Acad Emerg Med* 24(3): 328-336.

OBJECTIVES: The objective was to describe the epidemiology of dyspnea presenting to emergency departments (EDs) in the Asia-Pacific region, to understand how it is investigated and treated and its outcome. **METHODS:** Prospective interrupted time series cohort study conducted at three time points in EDs in Australia, New Zealand, Singapore, Hong Kong, and Malaysia of adult patients presenting to the ED with dyspnea as a main symptom. Data were collected over three 72-hour periods and included demographics, comorbidities, mode of arrival, usual medications, prehospital treatment, initial assessment, ED investigations, treatment in the ED, ED diagnosis, disposition from ED, in-hospital outcome, and final hospital diagnosis. The primary outcomes of interest are the epidemiology, investigation, treatment, and outcome of patients presenting to ED with dyspnea. **RESULTS:** A total of 3,044 patients

were studied. Patients with dyspnea made up 5.2% (3,105/60,059, 95% confidence interval [CI] = 5.0% to 5.4%) of ED presentations, 11.4% of ward admissions (1,956/17,184, 95% CI = 10.9% to 11.9%), and 19.9% of intensive care unit (ICU) admissions (104/523, 95% CI = 16.7% to 23.5%). The most common diagnoses were lower respiratory tract infection (20.2%), heart failure (14.9%), chronic obstructive pulmonary disease (13.6%), and asthma (12.7%). Hospital ward admission was required for 64% of patients (95% CI = 62% to 66%) with 3.3% (95% CI = 2.8% to 4.1%) requiring ICU admission. In-hospital mortality was 6% (95% CI = 5.0% to 7.2%). CONCLUSION: Dyspnea is a common symptom in ED patients contributing substantially to ED, hospital, and ICU workload. It is also associated with significant mortality. There are a wide variety of causes however chronic disease accounts for a large proportion.

Question: Disease Probability

Study Type: Prospective Cohort Study

Learner Level: Beginner

Notes: A study of disease probability for differential diagnosis in the clinical problem of dyspnea – this is a recent study of an important clinical problem that gets clinical learners' attention, and illustrates the need for diagnostic criteria and the challenges of applicability across countries.

Prandoni, P., et al. (2016). "Prevalence of Pulmonary Embolism among Patients Hospitalized for Syncope." *N Engl J Med* 375(16): 1524-1531.

BACKGROUND: The prevalence of pulmonary embolism among patients hospitalized for syncope is not well documented, and current guidelines pay little attention to a diagnostic workup for pulmonary embolism in these patients. **METHODS:** We performed a systematic workup for pulmonary embolism in patients admitted to 11 hospitals in Italy for a first episode of syncope, regardless of whether there were alternative explanations for the syncope. The diagnosis of pulmonary embolism was ruled out in patients who had a low pretest clinical probability, which was defined according to the Wells score, in combination with a negative d-dimer assay. In all other patients, computed tomographic pulmonary angiography or ventilation-perfusion lung scanning was performed. **RESULTS:** A total of 560 patients (mean age, 76 years) were included in the study. A diagnosis of pulmonary embolism was ruled out in 330 of the 560 patients (58.9%) on the basis of the combination of a low pretest clinical probability of pulmonary embolism and negative d-dimer assay. Among the remaining 230 patients, pulmonary embolism was identified in 97 (42.2%). In the entire cohort, the prevalence of pulmonary embolism was 17.3% (95% confidence interval, 14.2 to 20.5). Evidence of an embolus in a main pulmonary or lobar artery or evidence of perfusion defects larger than 25% of the total area of both lungs was found in 61 patients. Pulmonary embolism was identified in 45 of the 355 patients (12.7%) who had an alternative explanation for syncope and in 52 of the 205 patients (25.4%) who did not. **CONCLUSIONS:** Pulmonary embolism was identified in nearly one of every six patients hospitalized for a first episode of syncope. (Funded by the University of Padua; PESIT ClinicalTrials.gov number, NCT01797289).

Question: Disease Probability

Study Type: Prospective Cohort Study

Learner Level: Intermediate

Notes: Good in combination with article by Costantino on the same topic to compare and contrast methods of determining disease probability. The Costantino article demonstrates how data is obtained from administrative databases whereas the Prandoni paper is a prospective study with clinically recruited patients.

Qaseem, A., et al. (2006). "Risk assessment for and strategies to reduce perioperative pulmonary complications for patients undergoing noncardiothoracic surgery: a guideline from the American College of Physicians." *Ann Intern Med* 144(8): 575-580.

Postoperative pulmonary complications play an important role in the risk for patients undergoing noncardiothoracic surgery. Postoperative pulmonary complications are as prevalent as cardiac complications and contribute similarly to morbidity, mortality, and length of stay. Pulmonary complications may even be more likely than cardiac complications to predict long-term mortality after surgery. The purpose of this guideline is to provide guidance to clinicians on clinical and laboratory predictors of perioperative pulmonary risk before noncardiothoracic surgery. It also evaluates strategies to reduce the perioperative pulmonary risk and focuses on atelectasis, pneumonia, and respiratory failure. The target audience for this guideline is general internists or other clinicians involved in perioperative management of surgical patients. The target patient population is all adult persons undergoing noncardiothoracic surgery.

Question: Guideline

Study Type: Guideline

Learner Level: Intermediate

Notes: Guideline clearly reported from the ACP associated with a systematic review of the evidence published in the same journal of annals. This is a clearly reported guideline that is good for teaching as it is a manageable amount of information. (Many guidelines are difficult to teach because they are so comprehensive). Although the methods of the review are not clearly outlined in the text of the annals article, you can get a nice summary from the National Guidelines clearinghouse at www.guidelines.gov to use for teaching.

Prevention

Becattini, C., et al. (2012). "Aspirin for preventing the recurrence of venous thromboembolism." *N Engl J Med* 366(21): 1959-1967.

BACKGROUND: About 20% of patients with unprovoked venous thromboembolism have a recurrence within 2 years after the withdrawal of oral anticoagulant therapy. Extending anticoagulation prevents recurrences but is associated with increased bleeding. The benefit of aspirin for the prevention of recurrent venous thromboembolism is unknown. **METHODS:** In this multicenter, investigator-initiated, double-blind study, patients with first-ever unprovoked venous thromboembolism who had completed 6 to 18 months of oral anticoagulant treatment were randomly assigned to aspirin, 100 mg daily, or placebo for 2 years, with the option of extending the study treatment. The primary efficacy outcome was recurrence of venous thromboembolism, and major bleeding was the primary safety outcome. **RESULTS:** Venous thromboembolism recurred in 28 of the 205 patients who received aspirin and in 43 of the 197 patients who received placebo (6.6% vs. 11.2% per year; hazard ratio, 0.58; 95% confidence interval [CI], 0.36 to 0.93) (median study period, 24.6 months). During a median treatment period of 23.9 months, 23 patients taking aspirin and 39 taking placebo had a recurrence (5.9% vs. 11.0% per year; hazard ratio, 0.55; 95% CI, 0.33 to 0.92). One patient in each treatment group had a major bleeding episode. Adverse events were similar in the two groups. **CONCLUSIONS:** Aspirin reduced the risk of recurrence when given to patients with unprovoked venous thromboembolism who had discontinued anticoagulant treatment, with no apparent increase in the risk of major bleeding. (Funded by the University of Perugia and others; WARFASA ClinicalTrials.gov number, NCT00222677.)

Question: Prevention

Study Type: RCT

Learner Level: Beginner

Notes: WARFASA trial, positive trial of aspirin after stopping warfarin for VTE; a nice clear teaching example with good numbers for calculating ARR and RRR.

Bischoff-Ferrari, H. A., et al. (2009). "Fall prevention with supplemental and active forms of vitamin D: a meta-analysis of randomised controlled trials." *BMJ* 339: b3692.

OBJECTIVE: To test the efficacy of supplemental vitamin D and active forms of vitamin D with or without calcium in preventing falls among older individuals. **DATA SOURCES:** We searched Medline, the Cochrane central register of controlled trials, BIOSIS, and Embase up to August 2008 for relevant articles. Further studies were identified by consulting clinical experts, bibliographies, and abstracts. We contacted authors for additional data when necessary. **Review methods** Only double blind randomised controlled trials of older individuals (mean age 65 years or older) receiving a defined oral dose of supplemental vitamin D (vitamin D(3) (cholecalciferol) or vitamin D(2) (ergocalciferol)) or an active form of vitamin D (1alpha-hydroxyvitamin D(3) (1alpha-hydroxycalciferol) or 1,25-dihydroxyvitamin D(3) (1,25-dihydroxycholecalciferol)) and with sufficiently specified fall assessment were considered for inclusion. **RESULTS:** Eight randomised controlled trials (n=2426) of supplemental vitamin D met our inclusion criteria. Heterogeneity among trials was observed for dose of vitamin D (700-1000 IU/day v 200-600 IU/day; P=0.02) and achieved 25-hydroxyvitamin D(3) concentration (25(OH)D concentration: <60 nmol/l v >or=60 nmol/l; P=0.005). High dose supplemental vitamin D reduced fall risk by 19% (pooled relative risk (RR) 0.81, 95% CI 0.71 to 0.92; n=1921 from seven trials), whereas achieved serum 25(OH)D concentrations of 60 nmol/l or more resulted in a 23% fall reduction (pooled RR 0.77, 95% CI 0.65 to 0.90). Falls were not notably reduced by low dose supplemental vitamin D (pooled RR 1.10, 95% CI 0.89 to 1.35; n=505 from two trials) or by achieved serum 25-hydroxyvitamin D concentrations of less than 60 nmol/l (pooled RR 1.35, 95% CI 0.98 to 1.84). Two randomised controlled trials (n=624) of active forms of vitamin D met our inclusion criteria. Active forms of vitamin D reduced fall risk by 22% (pooled RR 0.78, 95% CI 0.64 to 0.94). **CONCLUSIONS:** Supplemental vitamin D in a dose of 700-1000 IU a day reduced the risk of falling among older individuals by 19% and to a similar degree as active forms of vitamin D. Doses of supplemental vitamin D of less than 700 IU or serum 25-hydroxyvitamin D concentrations of less than 60 nmol/l may not reduce the risk of falling among older individuals.

Question: Therapy

Study Type: Meta-analysis

Learner Level: Intermediate

Notes: Nicely reported meta-analysis that can be used to illustrate issues of heterogeneity. When all studies were included in the combined outcome, the results showed significant heterogeneity. However, if the data are separated into high dose and low dose, the heterogeneity is significantly improved. There is an ACP Journal club summary for this article.

Devereaux, P. J., et al. (2005). "How strong is the evidence for the use of perioperative beta blockers in non-cardiac surgery? Systematic review and meta-analysis of randomised controlled trials." *BMJ* 331(7512): 313-321.

OBJECTIVE: To determine the effect of perioperative beta blocker treatment in patients having non-cardiac surgery. **DESIGN:** Systematic review and meta-analysis. **DATA SOURCES:** Seven search strategies, including searching two bibliographic databases and hand searching seven medical journals. **STUDY SELECTION AND OUTCOMES:** We included randomised controlled trials that

evaluated beta blocker treatment in patients having non-cardiac surgery. Perioperative outcomes within 30 days of surgery included total mortality, cardiovascular mortality, non-fatal myocardial infarction, non-fatal cardiac arrest, non-fatal stroke, congestive heart failure, hypotension needing treatment, bradycardia needing treatment, and bronchospasm. RESULTS: Twenty two trials that randomised a total of 2437 patients met the eligibility criteria. Perioperative beta blockers did not show any statistically significant beneficial effects on any of the individual outcomes and the only nominally statistically significant beneficial relative risk was 0.44 (95% confidence interval 0.20 to 0.97, 99% confidence interval 0.16 to 1.24) for the composite outcome of cardiovascular mortality, non-fatal myocardial infarction, and non-fatal cardiac arrest. Methods adapted from formal interim monitoring boundaries applied to cumulative meta-analysis showed that the evidence failed, by a considerable degree, to meet standards for forgoing additional studies. The individual safety outcomes in patients treated with perioperative beta blockers showed a relative risk for bradycardia needing treatment of 2.27 (95% CI 1.53 to 3.36, 99% CI 1.36 to 3.80) and a nominally statistically significant relative risk for hypotension needing treatment of 1.27 (95% CI 1.04 to 1.56, 99% CI 0.97 to 1.66). CONCLUSION: The evidence that perioperative beta blockers reduce major cardiovascular events is encouraging but too unreliable to allow definitive conclusions to be drawn.

Question: Therapy

Study Type: Meta-analysis

Learner Level: Intermediate

Notes: Extremely rigorous methods clearly reported make this a great teaching paper. In addition, results are counter to current practice of broadly applying b-blockers in the perioperative period to patient undergoing non-cardiac surgery. Results are good for teaching about heterogeneity. The section in the results section on exploring heterogeneity can be very instructive.

Guay, L. A., et al. (1999). "Intrapartum and neonatal single-dose nevirapine compared with zidovudine for prevention of mother-to-child transmission of HIV-1 in Kampala, Uganda: HIVNET 012 randomised trial." *Lancet* 354(9181): 795-802.

BACKGROUND: The AIDS Clinical Trials Group protocol 076 zidovudine prophylaxis regimen for HIV-1-infected pregnant women and their babies has been associated with a significant decrease in vertical HIV-1 transmission in non-breastfeeding women in developed countries. We compared the safety and efficacy of short-course nevirapine or zidovudine during labour and the first week of life. METHODS: From November, 1997, to April, 1999, we enrolled 626 HIV-1-infected pregnant women at Mulago Hospital in Kampala, Uganda. We randomly assigned mothers nevirapine 200 mg orally at onset of labour and 2 mg/kg to babies within 72 h of birth, or zidovudine 600 mg orally to the mother at onset of labour and 300 mg every 3 h until delivery, and 4 mg/kg orally twice daily to babies for 7 days after birth. We tested babies for HIV-1 infection at birth, 6-8 weeks, and 14-16 weeks by HIV-1 RNA PCR. We assessed HIV-1 transmission and HIV-1-free survival with Kaplan-Meier analysis. FINDINGS: Nearly all babies (98.8%) were breastfed, and 95.6% were still breastfeeding at age 14-16 weeks. The estimated risks of HIV-1 transmission in the zidovudine and nevirapine groups were: 10.4% and 8.2% at birth ($p=0.354$); 21.3% and 11.9% by age 6-8 weeks ($p=0.0027$); and 25.1% and 13.1% by age 14-16 weeks ($p=0.0006$). The efficacy of nevirapine compared with zidovudine was 47% (95% CI 20-64) up to age 14-16 weeks. The two regimens were well tolerated and adverse events were similar in the two groups. INTERPRETATION: Nevirapine lowered the risk of HIV-1 transmission during the first 14-16 weeks of life by nearly 50% in a breastfeeding population. This simple and inexpensive regimen could decrease mother-to-child HIV-1 transmission in less-developed countries. PIP: A study was conducted to assess the safety and efficacy of short-course nevirapine compared with zidovudine given to women during labor and to neonates during the first week of life. 626 HIV-1 infected pregnant women attending the antenatal clinic from November 1997 to April 1999 at Mulago Hospital in Kampala, Uganda, were randomly given nevirapine or zidovudine. Infants were tested for HIV-1 infection at birth, at 6-8 weeks, and at 14-16 weeks. Findings revealed that the estimated risk of HIV-1 transmission in the zidovudine and nevirapine groups was 10.4% and 8.2%, respectively, at birth; 21.3% and 11.9%, by 6-8 weeks; and 25.1% and 13.1%, by 14-16 weeks. There was a 47% relative efficacy rate of the nevirapine regimen at 14-16 weeks compared to zidovudine. Based on the findings, nevirapine lowers the risk of HIV-1 transmission by nearly 50% during the first 14-16 weeks of life in breast-fed infants.

Question: Prevention

Study Type: RCT

Learner Level: Beginner / Intermediate

Notes: Strong methodology paper on prevention Good points for discussion: Good RCT methods, well described for outstanding discussion of issues pertaining to randomization, blinding (vs. open label) etc. Interesting discussion of how and whether the results can be generalized to other settings and potential social impact of this study.

Harper, C. C., et al. (2015). "Reductions in pregnancy rates in the USA with long-acting reversible contraception: a cluster randomised trial." *Lancet* 386(9993): 562-568.

BACKGROUND: Unintended pregnancy remains a serious public health challenge in the USA. We assessed the effects of an intervention to increase patients' access to long-acting reversible contraceptives (LARCs) on pregnancy rates. METHODS: We did a cluster randomised trial in 40 reproductive health clinics across the USA in 2011-13. 20 clinics were randomly assigned to receive evidence-based training on providing counselling and insertion of intrauterine devices (IUDs) or progestin implants and 20 to provide standard care. Usual costs for contraception were maintained at all sites. We recruited women aged 18-25 years attending family planning or abortion care visits and not desiring pregnancy in the next 12 months. The primary outcome was selection of an IUD or implant at the clinic visit and secondary outcome was pregnancy within 12 months. We used generalised estimating equations for clustered data to measure the intervention effect on contraceptive selection, and used survival analysis to assess pregnancy rates. FINDINGS: Of 1500 women enrolled, more at intervention than control sites reported receiving counselling on IUDs or implants (565

[71%] of 797 vs 271 [39%] of 693, odds ratio 3.8, 95% CI 2.8-5.2) and more selected LARCs during the clinic visit (224 [28%] vs 117 [17%], 1.9, 1.3-2.8). The pregnancy rate was lower in intervention group than in the control group after family planning visits (7.9 vs 15.4 per 100 person-years), but not after abortion visits (26.5 vs 22.3 per 100 person-years). We found a significant intervention effect on pregnancy rates in women attending family planning visits (hazard ratio 0.54, 95% CI 0.34-0.85). INTERPRETATION: The pregnancy rate can be reduced by provision of counselling on long-term reversible contraception and access to devices during family planning counselling visits. FUNDING: William and Flora Hewlett Foundation.

Question: Prevention

Study Type: RCT

Learner Level: Intermediate

Notes: About a health systems educational intervention using cluster randomization (Planned Parenthood clinics). Illustrates a variety of outcomes that can be measured with a systems educational intervention (patient knowledge, method selection, and pregnancy rates), and the role of cluster randomization in diminishing contamination. Other issues: blinding when using a behavioral intervention, and the use of sham interventions vs. no intervention as the comparator to behavioral interventions (no sham in this study.) Includes Kaplan-Meier curves. Can also be used to encourage learners to brainstorm on unintended consequences and generalizability when evaluating systems-level interventions.

Hayden, M., et al. (2002). "Aspirin for the primary prevention of cardiovascular events: a summary of the evidence for the U.S. Preventive Services Task Force." *Ann Intern Med* 136(2): 161-172.

BACKGROUND: The use of aspirin to prevent cardiovascular disease events in patients without a history of cardiovascular disease is controversial. **PURPOSE:** To examine the benefits and harms of aspirin chemoprevention. **DATA SOURCES:** MEDLINE (1966 to May 2001). **STUDY SELECTION:** 1) Randomized trials at least 1 year in duration that examined aspirin chemoprevention in patients without previously known cardiovascular disease and 2) systematic reviews, recent trials, and observational studies that examined rates of hemorrhagic strokes and gastrointestinal bleeding secondary to aspirin use. **DATA EXTRACTION:** One reviewer read and extracted data from each included article and constructed evidence tables. A second reviewer checked the accuracy of the data extraction. Discrepancies were resolved by consensus. **DATA SYNTHESIS:** Meta-analysis was performed, and the quantitative results of the review were then used to model the consequences of treating patients with different levels of baseline risk for coronary heart disease. Five trials examined the effect of aspirin on cardiovascular events in patients with no previous cardiovascular disease. For patients similar to those enrolled in the trials, aspirin reduces the risk for the combined end point of nonfatal myocardial infarction and fatal coronary heart disease (summary odds ratio, 0.72 [95% CI, 0.60 to 0.87]). Aspirin increased the risk for hemorrhagic strokes (summary odds ratio, 1.4 [CI, 0.9 to 2.0]) and major gastrointestinal bleeding (summary odds ratio, 1.7 [CI, 1.4 to 2.1]). All-cause mortality (summary odds ratio, 0.93 [CI, 0.84 to 1.02]) was not significantly affected. For 1000 patients with a 5% risk for coronary heart disease events over 5 years, aspirin would prevent 6 to 20 myocardial infarctions but would cause 0 to 2 hemorrhagic strokes and 2 to 4 major gastrointestinal bleeding events. For patients with a risk of 1% over 5 years, aspirin would prevent 1 to 4 myocardial infarctions but would cause 0 to 2 hemorrhagic strokes and 2 to 4 major gastrointestinal bleeding events. **CONCLUSIONS:** The net benefit of aspirin increases with increasing cardiovascular risk. In the decision to use aspirin chemoprevention, the patient's cardiovascular risk and relative utility for the different clinical outcomes prevented or caused by aspirin use must be considered.

Question: Prevention

Study Type: Practice guideline

Learner Level: Beginner / Intermediate

Notes: This pair of papers (see also Aspirin for... 2002) could be used as an example of the format and utility of the U.S. Preventive Services Task Force (USPSTF) Clinical Practice Guidelines. The Recommendation is brief but to the point in describing the recommendations and the strength of supporting evidence. When coupled with the Systematic Review of evidence companion article, learners can review the evidence in greater depth. Good discussion points: Excellent for discussion of weighing potential risks with potential benefits. Can include discussion of patient / societal values when trying to apply CPGs in any particular setting.

Kannus, P., et al. (2000). "Prevention of hip fracture in elderly people with use of a hip protector." *N Engl J Med* 343(21): 1506-1513.

BACKGROUND: Hip fractures are common in frail elderly adults worldwide. We investigated the effect of an anatomically designed external hip protector on the risk of these age-related fractures. **METHODS:** We randomly assigned 1801 ambulatory but frail elderly adults (1409 women and 392 men; mean age, 82 years), in a 1:2 ratio, either to a group that wore a hip protector or to a control group. Fractures of the hip and all other fractures were recorded until the end of the first full month after 62 hip fractures had occurred in the control group. The risk of fracture in the two groups was compared, and in the hip-protector group the risk of fracture was also analyzed according to whether the protector had been in use at the time of a fall. **RESULTS:** During follow-up, 13 subjects in the hip-protector group had a hip fracture, as compared with 67 subjects in the control group. The respective rates of hip fracture were 21.3 and 46.0 per 1000 person-years (relative hazard in the hip-protector group, 0.4; 95 percent confidence interval, 0.2 to 0.8; P=0.008). The risk of pelvic fracture was slightly but not significantly lower in the hip-protector group than in the control group (2 subjects and 12 subjects, respectively, had pelvic fracture) (relative hazard, 0.4; 95 percent confidence interval, 0.1 to 1.8; P > 0.05). The risk of other fractures was similar in the two groups. In the hip-protector group, four subjects had a hip fracture (among 1034 falls) while wearing the protector, and nine subjects had a hip fracture (among 370 falls) while not wearing the protector (relative hazard, 0.2; 95 percent confidence interval, 0.05 to 0.5; P=0.002). **CONCLUSIONS:** The risk of hip fracture can be reduced in frail elderly adults by the use of an anatomically designed external hip protector.

Question: Prevention

Study Type: RCT

Learner Level: Beginner / Intermediate

Notes: This is a strong, well done RCT that can be useful for teaching and learning about issues of prevention. Can also use to discuss hazard ratios Good paper for multidisciplinary discussion

Marenzi, G., et al. (2006). "N-acetylcysteine and contrast-induced nephropathy in primary angioplasty." *N Engl J Med* 354(26): 2773-2782.

BACKGROUND: Patients with acute myocardial infarction undergoing primary angioplasty are at high risk for contrast-medium-induced nephropathy because of hemodynamic instability, the need for a high volume of contrast medium, and the lack of effective prophylaxis. We investigated the antioxidant N-acetylcysteine for the prevention of contrast-medium-induced nephropathy in patients undergoing primary angioplasty. **METHODS:** We randomly assigned 354 consecutive patients undergoing primary angioplasty to one of three groups: 116 patients were assigned to a standard dose of N-acetylcysteine (a 600-mg intravenous bolus before primary angioplasty and 600 mg orally twice daily for the 48 hours after angioplasty), 119 patients to a double dose of N-acetylcysteine (a 1200-mg intravenous bolus and 1200 mg orally twice daily for the 48 hours after intervention), and 119 patients to placebo. **RESULTS:** The serum creatinine concentration increased 25 percent or more from baseline after primary angioplasty in 39 of the control patients (33 percent), 17 of the patients receiving standard-dose N-acetylcysteine (15 percent), and 10 patients receiving high-dose N-acetylcysteine (8 percent, $P<0.001$). Overall in-hospital mortality was higher in patients with contrast-medium-induced nephropathy than in those without such nephropathy (26 percent vs. 1 percent, $P<0.001$). Thirteen patients (11 percent) in the control group died, as did five (4 percent) in the standard-dose N-acetylcysteine group and three (3 percent) in the high-dose N-acetylcysteine group ($P=0.02$). The rate for the composite end point of death, acute renal failure requiring temporary renal-replacement therapy, or the need for mechanical ventilation was 21 (18 percent), 8 (7 percent), and 6 (5 percent) in the three groups, respectively ($P=0.002$). **CONCLUSIONS:** Intravenous and oral N-acetylcysteine may prevent contrast-medium-induced nephropathy with a dose-dependent effect in patients treated with primary angioplasty and may improve hospital outcome. (ClinicalTrials.gov number, NCT00237614[ClinicalTrials.gov]).

Question: Prevention

Study Type: RCT

Learner Level: Beginner to Intermediate

Notes: RCT that is clearly reported thus good for validity review. There are 3 arms (control, standard dose and high dose) to this randomized trial. The importance of this paper is in discussing surrogate outcomes (change in creatinine) versus 'patient important outcomes' (combined outcome of death, dialysis, mechanical ventilation). This study added an additional level of information to prior studies which focused only on surrogate (intermediate) outcomes. You can also discuss composite outcomes, their benefits and pitfalls.

Stampfer, M. J., et al. (1993). "Vitamin E consumption and the risk of coronary disease in women." *N Engl J Med* 328(20): 1444-1449.

BACKGROUND: Interest in the antioxidant vitamin E as a possible protective nutrient against coronary disease has intensified with the recognition that oxidized low-density lipoprotein may be involved in atherogenesis. **METHODS:** In 1980, 87,245 female nurses 34 to 59 years of age who were free of diagnosed cardiovascular disease and cancer completed dietary questionnaires that assessed their consumption of a wide range of nutrients, including vitamin E. During follow-up of up to eight years (679,485 person-years) that was 97 percent complete, we documented 552 cases of major coronary disease (437 nonfatal myocardial infarctions and 115 deaths due to coronary disease). **RESULTS:** As compared with women in the lowest fifth of the cohort with respect to vitamin E intake, those in the top fifth had a relative risk of major coronary disease of 0.66 (95 percent confidence interval, 0.50 to 0.87) after adjustment for age and smoking. Further adjustment for a variety of other coronary risk factors and nutrients, including other antioxidants, had little effect on the results. Most of the variability in intake and reduction in risk was attributable to vitamin E consumed as supplements. Women who took vitamin E supplements for short periods had little apparent benefit, but those who took them for more than two years had a relative risk of major coronary disease of 0.59 (95 percent confidence interval, 0.38 to 0.91) after adjustment for age, smoking status, risk factors for coronary disease, and use of other antioxidant nutrients (including multi-vitamins). **CONCLUSIONS:** Although these prospective data do not prove a cause-and-effect relation, they suggest that among middle-aged women the use of vitamin E supplements is associated with a reduced risk of coronary heart disease. Randomized trials of vitamin E in the primary and secondary prevention of coronary disease are being conducted; public policy recommendations about the widespread use of vitamin E should await the results of these trials.

Question: Prevention

Study Type: Cohort study

Learner Level: .

Notes: This is a great example of how a cohort study, even when well-done, cannot establish cause and effect. This very large prospective cohort study was well-done and meets just about all of the validity criteria for cohort studies. It seems to establish an

association between vitamin E consumption and reduced coronary disease. However, this was subsequently disproved on the basis of several well-done RCTs. So I use it to show that even a very well-done cohort CANNOT prove causation.

Turpie, A. G., et al. (2002). "Fondaparinux vs enoxaparin for the prevention of venous thromboembolism in major orthopedic surgery: a meta-analysis of 4 randomized double-blind studies." *Arch Intern Med* 162(16): 1833-1840.

BACKGROUND: Orthopedic surgery remains a condition at high risk of venous thromboembolism (VTE). Fondaparinux, the first of a new class of synthetic selective factor Xa inhibitors, may further reduce this risk compared with currently available thromboprophylactic treatments. **METHODS:** A meta-analysis of 4 multicenter, randomized, double-blind trials in patients undergoing elective hip replacement, elective major knee surgery, and surgery for hip fracture (N = 7344) was performed to determine whether a subcutaneous 2.5-mg, once-daily regimen of fondaparinux sodium starting 6 hours after surgery was more effective and as safe as approved enoxaparin regimens in preventing VTE. The primary efficacy outcome was VTE up to day 11, defined as deep vein thrombosis detected by mandatory bilateral venography or documented symptomatic deep vein thrombosis or pulmonary embolism. The primary safety outcome was major bleeding. **RESULTS:** Fondaparinux significantly reduced the incidence of VTE by day 11 (182 [6.8%] of 2682 patients) compared with enoxaparin (371 [13.7%] of 2703 patients), with a common odds reduction of 55.2% (95% confidence interval, 45.8% to 63.1%; P<.001); this beneficial effect was consistent across all types of surgery and all subgroups. Although major bleeding occurred more frequently in the fondaparinux-treated group (P = .008), the incidence of clinically relevant bleeding (leading to death or reoperation or occurring in a critical organ) did not differ between groups. **CONCLUSION:** In patients undergoing orthopedic surgery, 2.5 mg of fondaparinux sodium once daily, starting 6 hours postoperatively, showed a major benefit over enoxaparin, achieving an overall risk reduction of VTE greater than 50% without increasing the risk of clinically relevant bleeding.

Question: Therapy

Study Type: Meta-analysis

Learner Level: Beginner

Notes: Good to teach difference between a systematic review and a meta-analysis

Prognosis

Abramson, J. L., et al. (2001). "Moderate alcohol consumption and risk of heart failure among older persons." *JAMA* 285(15): 1971-1977.

CONTEXT: Heavy consumption of alcohol can lead to heart failure, but the relationship between moderate alcohol consumption and risk of heart failure is largely unknown. **OBJECTIVE:** To determine whether moderate alcohol consumption predicts heart failure risk among older persons, independent of the association of moderate alcohol consumption with lower risk of myocardial infarction (MI). **DESIGN:** Prospective cohort study conducted from 1982 through 1996, with a maximum follow-up of 14 years. **SETTING AND PARTICIPANTS:** Population-based sample of 2235 noninstitutionalized elderly persons (mean age, 73.7 years; 41.2% male; 21.3% nonwhite) residing in New Haven, Conn, who were free of heart failure at baseline. Persons who reported alcohol consumption of more than 70 oz in the month prior to baseline were excluded. **MAIN OUTCOME MEASURE:** Time to first fatal or nonfatal heart failure event, according to the amount of alcohol consumed in the month prior to baseline. **RESULTS:** Increasing alcohol consumption in the moderate range was associated with decreasing heart failure rates. For persons consuming no alcohol (50.0%), 1 to 20 oz (40.2%), and 21 to 70 oz (9.8%) in the month prior to baseline, crude heart failure rates per 1000 years of follow-up were 16.1, 12.2, and 9.2, respectively. After adjustment for age, sex, race, education, angina, history of MI and diabetes, MI during follow-up, hypertension, pulse pressure, body mass index, and current smoking, the relative risks of heart failure for those consuming no alcohol, 1 to 20 oz, and 21 to 70 oz in the month prior to baseline were 1.00 (referent), 0.79 (95% confidence interval [CI], 0.60-1.02), and 0.53 (95% CI, 0.32-0.88) (P for trend = .02). **CONCLUSIONS:** Increasing levels of moderate alcohol consumption are associated with a decreasing risk of heart failure among older persons. This association is independent of a number of confounding factors and does not appear to be entirely mediated by a reduction in MI risk.

Question: Prognosis

Study Type: Prospective Cohort

Learner Level: Beginner / Advanced

Notes: These two articles (see also Mukamal 2001) may be considered together (see sample case.) It is sometimes fun to compare and contrast two different articles about a similar prognosis topic. The populations studied through observational designs (in this case a cohort) are critical to how we understand and apply the information. Therefore, it is sometimes a good strategy to compare and contrast findings in studies done in different settings. In these two papers, there are many opportunities to discuss the impact of potential 'bias'; Try to consider why I chose these two papers- what about the two populations make for good comparison?? Note: either of these papers could also be looked at as a single paper.

Barefoot, J. C. and M. Schroll (1996). "Symptoms of depression, acute myocardial infarction, and total mortality in a community sample." *Circulation* 93(11): 1976-1980.

BACKGROUND: Depression has been shown to adversely affect the prognosis of patients with established coronary artery disease, but there is comparatively little evidence to document the role of depression in the initial development of coronary disease. **METHODS AND RESULTS:** Study participants were 409 men and 321 women who were residents of Glostrup, Denmark, born in 1914. Physical and psychological examinations in 1964 and 1974 established their baseline risk factor and disease status and their level of depressive symptomatology. Initial myocardial infarction (MI) was observed in 122 participants, and there were 290 deaths during follow-up, which ended in 1991. A 2-SD difference in depression score was associated with relative risks of 1.71 ($P = .005$) for MI and 1.59 ($P < .001$) for deaths from all causes. These findings were unchanged after we controlled for risk factors and signs of disease at baseline. There were no sex differences in effect sizes. **CONCLUSIONS:** High levels of depressive symptomatology are associated with increased risks of MI and mortality. The graded relationships between depression scores and risk, long-lasting nature of the effect, and stability of the depression measured across time suggest that this risk factor is best viewed as a continuous variable that represents a chronic psychological characteristic rather than a discrete and episodic psychiatric condition.

Question: Prognosis

Study Type: Prospective Cohort

Learner Level: Beginner / Intermediate

Notes: These two articles (see also Frasure-Smith 1993) may be considered together (see sample case.) It is sometimes fun to compare and contrast two different articles about a similar prognosis topic. The populations studied through observational designs (in this case a cohort) are critical to how we understand and apply the information. Therefore, it is sometimes a good strategy to compare and contrast findings in studies done in different settings. In these two papers, there are many opportunities to discuss the impact of potential 'bias'; Try to consider why I chose these two papers- what about the two populations make for good comparison?? Note: either of these papers could also be looked at as a single paper.

Butt, D. A., et al. (2013). "The risk of falls on initiation of antihypertensive drugs in the elderly." *Osteoporosis International* 24(10): 2649-2657.

SUMMARY: Antihypertensive drugs are associated with an immediate increased falls risk in elderly patients which was significant during the first 14 days after receiving a thiazide diuretic, angiotensin-converting enzyme inhibitor, angiotensin II receptor blocker, calcium channel blocker, or beta-adrenergic blocker. Fall prevention strategies during this period may prevent fall-related injuries. **INTRODUCTION:** The purpose of this study is to evaluate if initiation of the common antihypertensive drugs is associated with the occurrence of falls. **METHODS:** This population-based self-controlled case series study used healthcare administrative databases to identify new users of antihypertensive drugs in the elderly aged 66 and older living in Ontario, Canada who suffered a fall from April 1, 2000 to March 31, 2009. The risk period was the first 45 days following antihypertensive therapy initiation, further subdivided into 0-14 and 15-44 days with control periods before and after treatment in a 450-day observation period. We calculated the relative incidence (incidence rate ratio, IRR), defined as the rate of falls in the risk period compared to falls rate in the control periods. **RESULTS:** Of the 543,572 new users of antihypertensive drugs among community-dwelling elderly, 8,893 experienced an injurious fall that required hospital care during the observation period. New users had a 69% increased risk of having an injurious fall during the first 45 days following antihypertensive treatment (IRR = 1.69; 95% CI, 1.57-1.81). This finding was consistent for thiazide diuretics, angiotensin-converting enzyme inhibitors, calcium channel blockers, and beta-adrenergic blockers but not angiotensin II receptor antagonists. There was also an increased falls risk during the first 14 days of antihypertensive drug initiation (IRR = 1.94; 95% CI, 1.75-2.16), which was consistent for all antihypertensive drug classes. **CONCLUSIONS:** This study suggests that initiation of antihypertensive drugs is a risk factor for falls in the elderly. Fall prevention strategies during this period may reduce injuries.

Question: Harm

Study Type: Case Series

Learner Level: Advanced

Notes: Prognosis/Harm – more challenging but interesting article for advanced EBMers on risk of falls at various time periods following initiation of antihypertensive medications, using self-controlled case series design so each participant serves as their own control.

Chandratheva, A., et al. (2009). "Population-based study of risk and predictors of stroke in the first few hours after a TIA." *Neurology* 72(22): 1941-1947.

BACKGROUND: Several recent guidelines recommend assessment of patients with TIA within 24 hours, but it is uncertain how many recurrent strokes occur within 24 hours. It is also unclear whether the ABCD2 risk score reliably identifies recurrences in the first few hours. **METHODS:** In a prospective, population-based incidence study of TIA and stroke with complete follow-up (Oxford Vascular Study), we determined the 6-, 12-, and 24-hour risks of recurrent stroke, defined as new neurologic symptoms of sudden onset after initial recovery. **RESULTS:** Of 1,247 first TIA or strokes, 35 had recurrent strokes within 24 hours, all in the same arterial territory. The initial event had recovered prior to the recurrent stroke (i.e., was a TIA) in 25 cases. The 6-, 12-, and 24-hour stroke risks after 488 first TIAs were 1.2% (95% confidence interval [CI]: 0.2-2.2), 2.1% (0.8-3.2), and 5.1% (3.1-7.1), with 42% of all strokes during the 30 days after a first TIA occurring within the first 24 hours. The 12- and 24-hour risks were strongly related to ABCD2 score ($p = 0.02$ and $p = 0.0003$). Sixteen (64%) of the 25 cases sought urgent medical attention prior to the recurrent stroke, but none received antiplatelet treatment acutely. **CONCLUSION:** That about half of all recurrent strokes during the 7 days after a TIA occur in the first 24 hours highlights the need for emergency assessment. That the ABCD2 score is reliable in the hyperacute phase shows that appropriately triaged emergency assessment and treatment are feasible.

Question: Prognosis

Study Type: Prospective Cohort

Learner Level: Beginner

Notes: very nicely done, covers all the elements for critical appraisal of a prognosis paper

Condat, B., et al. (2001). "Current outcome of portal vein thrombosis in adults: risk and benefit of anticoagulant therapy." *Gastroenterology* 120(2): 490-497.

BACKGROUND & AIMS: The outcome of portal vein thrombosis in relation to associated prothrombotic states has not been evaluated. We assessed current outcome and predictors of bleeding and thrombotic events in a cohort of 136 adults with nonmalignant, noncirrhotic portal vein thrombosis, of whom 84 received anticoagulant therapy. **METHODS:** Multivariate Cox model analysis for event-free survival and analysis taking into account multiple events were used. **RESULTS:** Median follow-up was 46 months. The incidence rate of gastrointestinal bleeding was 12.5 (95% confidence interval [CI], 10-15) per 100 patient-years. Large varices were an independent predictor for bleeding. Anticoagulant therapy did not increase the risk or the severity of bleeding. The incidence rate of thrombotic events was 5.5 (95% CI, 3.8-7.2) per 100 patient-years. Underlying prothrombotic state and absence of anticoagulant therapy were independent predictors for thrombosis. In patients with underlying prothrombotic state, the incidence rates of splanchnic venous infarction were 0.82 and 5.2 per 100 patient-years in periods with and without anticoagulant therapy, respectively ($P = 0.01$). Two nonanticoagulated patients died of bleeding and thrombosis, respectively. **CONCLUSIONS:** In patients with portal vein thrombosis, the risk of thrombosis is currently as clinically significant as the risk of bleeding. The benefit-risk ratio favors anticoagulant therapy.

Question: Prognosis

Study Type: Cohort Study

Learner Level: Intermediate - Advanced

Notes: Cohort study that records natural history and prognosis; Also tries to make inferences about interventions which is much more tricky, but in scenarios where there may not be RCT data.

Correll, C. U., et al. (2009). "Cardiometabolic risk of second-generation antipsychotic medications during first-time use in children and adolescents." *JAMA* 302(16): 1765-1773.

CONTEXT: Cardiometabolic effects of second-generation antipsychotic medications are concerning but have not been sufficiently studied in pediatric and adolescent patients naive to antipsychotic medication. **OBJECTIVE:** To study the association of second-generation antipsychotic medications with body composition and metabolic parameters in patients without prior antipsychotic medication exposure. **DESIGN, SETTING, AND PATIENTS:** Nonrandomized Second-Generation Antipsychotic Treatment Indications, Effectiveness and Tolerability in Youth (SATIETY) cohort study, conducted between December 2001 and September 2007 at semi-urban, tertiary care, academic inpatient and outpatient clinics in Queens, New York, with a catchment area of 4.5-million individuals. Of 505 youth aged 4 to 19 years with 1 week or less of antipsychotic medication exposure, 338 were enrolled (66.9%). Of these patients, 272 had at least 1 postbaseline assessment (80.5%), and 205 patients [corrected] completed the study (60.7%). Patients had mood spectrum ($n = 130$; 47.8%), schizophrenia spectrum ($n = 82$; 30.1%), and disruptive or aggressive behavior spectrum ($n = 60$; 22.1%) disorders. Fifteen patients who refused participation or were nonadherent served as a comparison group. **INTERVENTION:** Treatment with aripiprazole, olanzapine, quetiapine, or risperidone for 12 weeks. **MAIN OUTCOME MEASURES:** Weight gain and changes in lipid and metabolic parameters. **RESULTS:** After a median of 10.8 weeks (interquartile range, 10.5-11.2 weeks) of treatment, weight increased by 8.5 kg (95% confidence interval [CI], 7.4 to 9.7 kg) with olanzapine ($n = 45$), by 6.1 kg (95% CI, 4.9 to 7.2 kg) with quetiapine ($n = 36$), by 5.3 kg (95% CI, 4.8 to 5.9 kg) with risperidone ($n = 135$), and by 4.4 kg (95% CI, 3.7 to 5.2 kg) with aripiprazole ($n = 41$) compared with the minimal weight change of 0.2 kg (95% CI, -1.0 to 1.4 kg) in the untreated comparison group ($n = 15$). With olanzapine and quetiapine, respectively, mean levels increased significantly for total cholesterol (15.6 mg/dL [95% CI, 6.9 to 24.3 mg/dL] $P < .001$ and 9.1 mg/dL [95% CI, 0.4 to 17.7 mg/dL] $P = .046$), triglycerides (24.3 mg/dL [95% CI, 9.8 to 38.9 mg/dL] $P = .002$ and 37.0 mg/dL [95% CI, 10.1 to 63.8 mg/dL] $P = .01$), non-high-density lipoprotein (HDL) cholesterol (16.8 mg/dL [95% CI, 9.3 to 24.3 mg/dL] $P < .001$ and 9.9 mg/dL [95% CI, 1.4 to 18.4 mg/dL] $P = .03$), and ratio of triglycerides to HDL cholesterol (0.6 [95% CI, 0.2 to 0.9] $P = .002$ and 1.2 [95% CI, 0.4 to 2.0] $P = .004$). With risperidone, triglycerides increased significantly (mean level, 9.7 mg/dL [95% CI, 0.5 to 19.0 mg/dL]; $P = .04$). Metabolic baseline-to-end-point changes were not significant with aripiprazole or in the untreated comparison group. **CONCLUSIONS:** First-time second-generation antipsychotic medication use was associated with significant weight gain with each medication. Metabolic changes varied among the 4 antipsychotic medications.

Question: Prognosis

Study Type: Prospective Cohort

Learner Level: Intermediate to Advanced

Notes: This cohort study examines the association of antipsychotic medications with weight gain and metabolic parameters in children and adolescents. As medication usage in children and adolescents continues to increase, the study of consequences of this

use are increasingly important. This study can be used to discuss the relationship between prognosis (the distribution of outcomes over time) and harm. Learners may need to be guided to the correct critical appraisal sheet for a cohort study (prognosis sheet).

Eisenberg, M. A., et al. (2013). "Time interval between concussions and symptom duration." *Pediatrics* 132(1): 8-17.

OBJECTIVE: To test the hypothesis that children with a previous history of concussion have a longer duration of symptoms after a repeat concussion than those without such a history. **METHODS:** Prospective cohort study of consecutive patients 11 to 22 years old presenting to the emergency department of a children's hospital with an acute concussion. The main outcome measure was time to symptom resolution, assessed by the Rivermead Post-Concussion Symptoms Questionnaire (RPSQ). Patients and providers completed a questionnaire describing mechanism of injury, associated symptoms, past medical history, examination findings, diagnostic studies, and the RPSQ. Patients were then serially administered the RPSQ for 3 months after the concussion or until all symptoms resolved. **RESULTS:** A total of 280 patients were enrolled over 12 months. Patients with a history of previous concussion had a longer duration of symptoms than those without previous concussion (24 vs 12 days, $P = .02$). Median symptom duration was even longer for patients with multiple previous concussions (28 days, $P = .03$) and for those who had sustained a concussion within the previous year (35 days, $P = .007$) compared with patients without those risk factors. In a multivariate model, previous concussion, absence of loss of consciousness, age ≥ 13 , and initial RPSQ score >18 were significant predictors of prolonged recovery. **CONCLUSIONS:** Children with a history of a previous concussion, particularly recent or multiple concussions, are at increased risk for prolonged symptoms after concussion. These findings have direct implications on the management of patients with concussion who are at high risk for repeat injuries.

Question: Prognosis

Study Type: Prospective cohort study

Learner Level: Beginner

Notes: Great clinical application and the use of survival curves on improvement in outcomes (rather than death).

El-Solh, A. A., et al. (2003). "Microbiology of severe aspiration pneumonia in institutionalized elderly." *Am J Respir Crit Care Med* 167(12): 1650-1654.

We sought to investigate prospectively the microbial etiology and prognostic indicators of 95 institutionalized elders with severe aspiration pneumonia, and to investigate its relation to oral hygiene in using quantitative bronchial sampling. Data collection included demographic information, Activity of Daily Living, Plaque Index, antimicrobial therapy, and outcome. Out of the 67 pathogens identified, Gram-negative enteric bacilli were the predominant organisms isolated (49%), followed by anaerobic bacteria (16%), and *Staphylococcus aureus* (12%). The most commonly encountered anaerobes were *Prevotella* and *Fusobacterium* species. Aerobic Gram-negative bacilli were recovered in conjunction with 55% of anaerobic isolates. The Plaque Index did not differ significantly between the aerobic (2.2 +/- 0.4) and the anaerobic group (2.3 +/- 0.3). Functional status was the only determinant of the presence of anaerobic bacteria. Although seven cases with anaerobic isolates received initially inadequate antimicrobial therapy, six had effective clinical response. The crude mortality was 33% for the aerobic and 36% for the anaerobic group ($p = 0.9$). Stepwise multivariate analysis identified hypoalbuminemia ($p < 0.001$) and the burden of comorbid diseases ($p < 0.001$) as independent risk factors of poor outcome. In view of the rising resistance to antimicrobial agents, the importance of adding anaerobic coverage for aspiration pneumonia in institutionalized elders needs to be reexamined.

Question: Prognosis

Study Type: cohort study

Learner Level: Beginner

Notes: Prospective cohort study looking at etiology and also prognosis; Clear methods and descriptive distribution of microbiologic diagnoses; Also includes prognosis information including length of intubation, ICU stay and mortality

Fernandez Perez, E. R., et al. (2013). "Identifying an inciting antigen is associated with improved survival in patients with chronic hypersensitivity pneumonitis." *Chest* 144(5): 1644-1651.

BACKGROUND: The cornerstone of hypersensitivity pneumonitis (HP) management is having patients avoid the inciting antigen (IA). Often, despite an exhaustive search, an IA cannot be found. The objective of this study was to examine whether identifying the IA impacts survival in patients with chronic HP. **METHODS:** We used the Kaplan-Meier method to display, and the log-rank test to compare, survival curves of patients with well-characterized chronic HP stratified on identification of an IA exposure. A Cox proportional hazards (PH) model was used to identify independent predictors in time-to-death analysis. **RESULTS:** Of 142 patients, 67 (47%) had an identified IA, and 75 (53%) had an unidentified IA. Compared with survivors, patients who died ($n = 80$, 56%) were older, more likely to have smoked, had lower total lung capacity % predicted and FVC % predicted, had higher severity of dyspnea, were more likely to have pulmonary fibrosis, and were less likely to have an identifiable IA. In a Cox PH model, the inability to identify an IA (hazard ratio [HR], 1.76; 95% CI, 1.01-3.07), older age (HR, 1.04; 95% CI, 1.01-1.07), the presences of pulmonary fibrosis (HR, 2.43; 95% CI, 1.36-4.35), a lower FVC% (HR, 1.36; 95% CI, 1.10-1.68), and a history of smoking (HR, 2.01; 95% CI, 1.15-3.50) were independent predictors of shorter survival. After adjusting for mean age, presence of fibrosis, mean FVC%, mean diffusing capacity of the lung for carbon monoxide (%), and history of smoking, survival was longer for patients with an identified IA exposure than those with an unidentified IA exposure (median, 8.75 years vs 4.88 years; $P = .047$). **CONCLUSIONS:** Among patients

with chronic HP, when adjusting for a number of potentially influential predictors, including the presence of fibrosis, the inability to identify an IA was independently associated with shortened survival.

Question: Prognosis

Study Type: Retrospective Cohort Study

Learner Level: Beginner

Notes: Good for looking at hazard ratios and survival curves.

Fine, M. J., et al. (1997). "A prediction rule to identify low-risk patients with community-acquired pneumonia." *N Engl J Med* 336(4): 243-250.

BACKGROUND: There is considerable variability in rates of hospitalization of patients with community-acquired pneumonia, in part because of physicians' uncertainty in assessing the severity of illness at presentation. **METHODS:** From our analysis of data on 14,199 adult inpatients with community-acquired pneumonia, we derived a prediction rule that stratifies patients into five classes with respect to the risk of death within 30 days. The rule was validated with 1991 data on 38,039 inpatients and with data on 2287 inpatients and outpatients in the Pneumonia Patient Outcomes Research Team (PORT) cohort study. The prediction rule assigns points based on age and the presence of coexisting disease, abnormal physical findings (such as a respiratory rate of $> \text{ or } = 30$ or a temperature of $> \text{ or } = 40$ degrees C), and abnormal laboratory findings (such as a pH < 7.35 , a blood urea nitrogen concentration $> \text{ or } = 30$ mg per deciliter [$11 \text{ mmol per liter}$] or a sodium concentration $< 130 \text{ mmol per liter}$) at presentation. **RESULTS:** There were no significant differences in mortality in each of the five risk classes among the three cohorts. Mortality ranged from 0.1 to 0.4 percent for class I patients ($P=0.22$), from 0.6 to 0.7 percent for class II ($P=0.67$), and from 0.9 to 2.8 percent for class III ($P=0.12$). Among the 1575 patients in the three lowest risk classes in the Pneumonia PORT cohort, there were only seven deaths, of which only four were pneumonia-related. The risk class was significantly associated with the risk of subsequent hospitalization among those treated as outpatients and with the use of intensive care and the number of days in the hospital among inpatients. **CONCLUSIONS:** The prediction rule we describe accurately identifies the patients with community-acquired pneumonia who are at low risk for death and other adverse outcomes. This prediction rule may help physicians make more rational decisions about hospitalization for patients with pneumonia.

Question: Prognosis

Study Type: Clinical prediction rule

Learner Level: Advanced

Notes: This is an example of a clinical prediction rule (a sum of factors used as a group to predict an outcome). This paper has a good example discussion of a derivation set (data used to develop the model to begin with) as compared with a validation set (once developed, different data is used to 'check' if the model makes sense.) This is a difficult paper for someone with more considerable experience looking for a challenge.

Frasure-Smith, N., et al. (1993). "Depression following myocardial infarction. Impact on 6-month survival." *JAMA* 270(15): 1819-1825.

OBJECTIVE--To determine if the diagnosis of major depression in patients hospitalized following myocardial infarction (MI) would have an independent impact on cardiac mortality over the first 6 months after discharge. **DESIGN--**Prospective evaluation of the impact of depression assessed using a modified version of the National Institute of Mental Health Diagnostic Interview Schedule for major depressive episode. Cox proportional hazards regression was used to evaluate the independent impact of depression after control for significant clinical predictors in the data set. **SETTING--**A large, university-affiliated hospital specializing in cardiac care, located in Montreal, Quebec. **PATIENTS--**All consenting patients ($N = 222$) who met established criteria for MI between August 1991 and July 1992 and who survived to be discharged from the hospital. Patients were interviewed between 5 and 15 days following the MI and were followed up for 6 months. There were no age limits (range, 24 to 88 years; mean, 60 years). The sample was 78% male. **PRIMARY OUTCOME MEASURE--**Survival status at 6 months. **RESULTS--**By 6 months, 12 patients had died. All deaths were due to cardiac causes. Depression was a significant predictor of mortality (hazard ratio, 5.74; 95% confidence interval, 4.61 to 6.87; $P = .0006$). The impact of depression remained after control for left ventricular dysfunction (Killip class) and previous MI, the multivariate significant predictors of mortality in the data set (adjusted hazard ratio, 4.29; 95% confidence interval, 3.14 to 5.44; $P = .013$). **CONCLUSION--**Major depression in patients hospitalized following an MI is an independent risk factor for mortality at 6 months. Its impact is at least equivalent to that of left ventricular dysfunction (Killip class) and history of previous MI. Additional study is needed to determine whether treatment of depression can influence post-MI survival and to assess possible underlying mechanisms.

Question: Prognosis

Study Type: Prospective Cohort

Learner Level: Beginner / Intermediate

Notes: These two articles (see also Barefoot 1996) may be considered together (see sample case.) It is sometimes fun to compare and contrast two different articles about a similar prognosis topic. The populations studied through observational designs (in this case a cohort) are critical to how we understand and apply the information. Therefore, it is sometimes a good strategy to compare and contrast findings in studies done in different settings. In these two papers, there are many opportunities to discuss the impact of

potential 'bias'; Try to consider why I chose these two papers- what about the two populations make for good comparison?? Note: either of these papers could also be looked at as a single paper.

Fukuda, T., et al. (2016). "Conventional Versus Compression-Only Versus No-Bystander Cardiopulmonary Resuscitation for Pediatric Out-of-Hospital Cardiac Arrest." *Circulation* 134(25): 2060-2070.

BACKGROUND: Conventional cardiopulmonary resuscitation (CPR) (chest compression and rescue breathing) has been recommended for pediatric out-of-hospital cardiac arrest (OHCA) because of the asphyxial nature of the majority of pediatric cardiac arrest events. However, the clinical effectiveness of additional rescue breathing (conventional CPR) compared with compression-only CPR in children is uncertain. **METHODS:** This nationwide population-based study of pediatric OHCA patients was based on data from the All-Japan Utstein Registry. We included all pediatric patients who experienced OHCA in Japan from January 1, 2011, to December 31, 2012. The primary outcome was a favorable neurological state 1 month after OHCA defined as a Glasgow-Pittsburgh Cerebral Performance Category score of 1 to 2 (corresponding to a Pediatric Cerebral Performance Category score of 1-3). Outcomes were compared with logistic regression with uni- and multivariable modeling in the overall cohort and for a propensity-matched subset of patients. **RESULTS:** A total of 2157 patients were included; 417 received conventional CPR, 733 received compression-only CPR, and 1007 did not receive any bystander CPR. Among these patients, 213 (9.9%) survived with a favorable neurological status 1 month after OHCA, including 108/417 (25.9%) for conventional, 68/733 (9.3%) for compression-only, and 37/1007 (3.7%) for no-bystander CPR. In unadjusted analyses, conventional CPR was superior to compression-only CPR in neurologically favorable survival (odds ratio [OR] 3.42, 95% confidence interval [CI] 2.45-4.76; $P < 0.0001$), with a trend favoring conventional CPR that was no longer statistically significant after multivariable adjustment (ORadjusted 1.52, 95% CI 0.93-2.49), and with further attenuation of the difference in a propensity-matched subset (OR 1.20, 95% CI 0.81-1.77). Both conventional and compression-only CPR were associated with higher odds for neurologically favorable survival compared with no-bystander CPR (ORadjusted 5.01, 95% CI 2.98-8.57, and ORadjusted 3.29, 95% CI 1.93-5.71), respectively. **CONCLUSIONS:** In this population-based study of pediatric OHCA in Japan, both conventional and compression-only CPR were associated with superior outcomes compared with no-bystander CPR. Unadjusted outcomes with conventional CPR were superior to compression-only CPR, with the magnitude of difference attenuated and no longer statistically significant after statistical adjustments. These findings support randomized clinical trials comparing conventional versus compression-only CPR in children, with conventional CPR preferred until such controlled comparative data are available, and either method preferred over no-bystander CPR.

Question: Prognosis

Study Type: .

Learner Level: Beginner/Intermediate

Notes: Does well with meeting validity criteria. Robust methodology overall. Very large sample size for this type of clinical question. Output for some outcomes is odds ratios so can teach on that. Can also have a nice discussion on outcomes reported – clinically important? Patient-centered?

Gage, B. F., et al. (2001). "Validation of clinical classification schemes for predicting stroke: results from the National Registry of Atrial Fibrillation." *JAMA* 285(22): 2864-2870.

CONTEXT: Patients who have atrial fibrillation (AF) have an increased risk of stroke, but their absolute rate of stroke depends on age and comorbid conditions. **OBJECTIVE:** To assess the predictive value of classification schemes that estimate stroke risk in patients with AF. **DESIGN, SETTING, AND PATIENTS:** Two existing classification schemes were combined into a new stroke-risk scheme, the CHADS(2) index, and all 3 classification schemes were validated. The CHADS(2) was formed by assigning 1 point each for the presence of congestive heart failure, hypertension, age 75 years or older, and diabetes mellitus and by assigning 2 points for history of stroke or transient ischemic attack. Data from peer review organizations representing 7 states were used to assemble a National Registry of AF (NRAF) consisting of 1733 Medicare beneficiaries aged 65 to 95 years who had nonrheumatic AF and were not prescribed warfarin at hospital discharge. **MAIN OUTCOME MEASURE:** Hospitalization for ischemic stroke, determined by Medicare claims data. **RESULTS:** During 2121 patient-years of follow-up, 94 patients were readmitted to the hospital for ischemic stroke (stroke rate, 4.4 per 100 patient-years). As indicated by a c statistic greater than 0.5, the 2 existing classification schemes predicted stroke better than chance: c of 0.68 (95% confidence interval [CI], 0.65-0.71) for the scheme developed by the Atrial Fibrillation Investigators (AFI) and c of 0.74 (95% CI, 0.71-0.76) for the Stroke Prevention in Atrial Fibrillation (SPAF) III scheme. However, with a c statistic of 0.82 (95% CI, 0.80-0.84), the CHADS(2) index was the most accurate predictor of stroke. The stroke rate per 100 patient-years without antithrombotic therapy increased by a factor of 1.5 (95% CI, 1.3-1.7) for each 1-point increase in the CHADS(2) score: 1.9 (95% CI, 1.2-3.0) for a score of 0; 2.8 (95% CI, 2.0-3.8) for 1; 4.0 (95% CI, 3.1-5.1) for 2; 5.9 (95% CI, 4.6-7.3) for 3; 8.5 (95% CI, 6.3-11.1) for 4; 12.5 (95% CI, 8.2-17.5) for 5; and 18.2 (95% CI, 10.5-27.4) for 6. **CONCLUSION:** The 2 existing classification schemes and especially a new stroke risk index, CHADS(2), can quantify risk of stroke for patients who have AF and may aid in selection of antithrombotic therapy.

Question: Prognosis

Study Type: Clinical prediction Rule

Learner Level: Intermediate to Advanced

Notes: Useful prediction rule that utilizes data from multiple large studies to compare schemes for predicting risk of stroke in atrial fibrillation. While the methods may be difficult for less experienced learners, the applicability of this tool makes for compelling discussion. Also, a general understanding of clinical decision rules can be put forward.

Hasbun, R., et al. (2001). "Computed tomography of the head before lumbar puncture in adults with suspected meningitis." *N Engl J Med* 345(24): 1727-1733.

BACKGROUND: In adults with suspected meningitis clinicians routinely order computed tomography (CT) of the head before performing a lumbar puncture. **METHODS:** We prospectively studied 301 adults with suspected meningitis to determine whether clinical characteristics that were present before CT of the head was performed could be used to identify patients who were unlikely to have abnormalities on CT. The Modified National Institutes of Health Stroke Scale was used to identify neurologic abnormalities. **RESULTS:** Of the 301 patients with suspected meningitis, 235 (78 percent) underwent CT of the head before undergoing lumbar puncture. In 56 of the 235 patients (24 percent), the results of CT were abnormal; 11 patients (5 percent) had evidence of a mass effect. The clinical features at base line that were associated with an abnormal finding on CT of the head were an age of at least 60 years, immunocompromise, a history of central nervous system disease, and a history of seizure within one week before presentation, as well as the following neurologic abnormalities: an abnormal level of consciousness, an inability to answer two consecutive questions correctly or to follow two consecutive commands, gaze palsy, abnormal visual fields, facial palsy, arm drift, leg drift, and abnormal language (e.g., aphasia). None of these features were present at base line in 96 of the 235 patients who underwent CT scanning of the head (41 percent). The CT scan was normal in 93 of these 96 patients, yielding a negative predictive value of 97 percent. Of the three misclassified patients, only one had a mild mass effect on CT, and all three subsequently underwent lumbar puncture, with no evidence of brain herniation one week later. **CONCLUSIONS:** In adults with suspected meningitis, clinical features can be used to identify those who are unlikely to have abnormal findings on CT of the head.

Question: Prognosis

Study Type: Prospective Cohort

Learner Level: Intermediate

Notes: Solid methodology, clearly reported. This prospective cohort study is an excellent example of how associations can be used in clinical decision making. Of note, you can also touch on principles of diagnosis/clinical decision rules (figure 5) when the authors test the diagnostic utility of combinations of baseline characteristics to predict normal or abnormal head CT.

Helgason, S., et al. (2000). "Prevalence of postherpetic neuralgia after a first episode of herpes zoster: prospective study with long term follow up." *BMJ* 321(7264): 794-796.

OBJECTIVE: To estimate the frequency, duration, and clinical importance of postherpetic neuralgia after a single episode of herpes zoster. **DESIGN:** Prospective cohort study with long term follow up. **SETTING:** Primary health care in Iceland. **PARTICIPANTS:** 421 patients with a single episode of herpes zoster. **MAIN OUTCOME MEASURES:** Age and sex distribution of patients with herpes zoster, point prevalence of postherpetic neuralgia, and severity of pain at 1, 3, 6, and 12 months and up to 7.6 years after the outbreak of zoster. **RESULTS:** Among patients younger than 60 years, the risk of postherpetic neuralgia three months after the start of the zoster rash was 1.8% (95% confidence interval 0.59% to 4.18%) and pain was mild in all cases. In patients 60 years and older, the risk of postherpetic neuralgia increased but the pain was usually mild or moderate. After three months severe pain was recorded in two patients older than 60 years (1.7%, 2.14% to 6.15%). After 12 months no patient reported severe pain and 14 patients (3.3%) had mild or moderate pain. Seven of these became pain free within two to seven years, and five reported mild pain and one moderate pain after 7.6 years of follow up. Sex was not a predictor of postherpetic neuralgia. Possible immunomodulating comorbidity (such as malignancy, systemic steroid use, diabetes) was present in 17 patients. **CONCLUSIONS:** The probability of longstanding pain of clinical importance after herpes zoster is low in an unselected population of primary care patients essentially untreated with antiviral drugs.

Question: Prognosis

Study Type: Prospective Cohort

Learner Level: Beginner

Notes: Straightforward methods for a prospective cohort trial with long term follow up. Good discussion points: Can discuss Kaplan Meier plot, prevalence, odds ratios and confidence intervals

Hoberman, A., et al. (2003). "Imaging studies after a first febrile urinary tract infection in young children." *N Engl J Med* 348(3): 195-202.

BACKGROUND: Guidelines from the American Academy of Pediatrics recommend obtaining a voiding cystourethrogram and a renal ultrasonogram for young children after a first urinary tract infection; renal scanning with technetium-99m-labeled dimercaptosuccinic acid has also been endorsed by other authorities. We investigated whether imaging studies altered management or improved outcomes in young children with a first febrile urinary tract infection. **METHODS:** In a prospective trial involving 309 children (1 to 24 months old), an ultrasonogram and an initial renal scan were obtained within 72 hours after diagnosis, contrast voiding cystourethrogram was performed one month later, and renal scanning was repeated six months later. **RESULTS:** The ultrasonographic results were normal in 88 percent of the children (272 of 309); the identified abnormalities did not modify management. Acute pyelonephritis was diagnosed in 61 percent of the children (190 of 309). Thirty-nine percent of the children who underwent cystourethrogram (117 of 302) had vesicoureteral reflux; 96 percent of these children (112 of 117) had grade I, II, or III vesicoureteral reflux. Repeated scans were obtained for 89 percent of the children (275 of 309); renal scarring was noted in 9.5 percent of these children (26 of 275). **CONCLUSIONS:** An ultrasonogram performed at the time of acute illness is of limited value. A voiding cystourethrogram for the identification of reflux is useful only if antimicrobial prophylaxis is effective in reducing reinfections

and renal scarring. Renal scans obtained at presentation identify children with acute pyelonephritis, and scans obtained six months later identify those with renal scarring. The routine performance of urinalysis, urine culture, or both during subsequent febrile illnesses in all children with a previous febrile urinary tract infection will probably obviate the need to obtain either early or late scans.

Question: Prognosis

Study Type: Prospective Cohort

Learner Level: Beginner / Advanced

Notes: Prospective cohort in children. This paper can be used to trigger discussion of prognosis as well as diagnosis.

Johnston, S. C., et al. (2000). "Short-term prognosis after emergency department diagnosis of TIA." *JAMA* 284(22): 2901-2906.

CONTEXT: Management of patients with acute transient ischemic attack (TIA) varies widely, with some institutions admitting all patients and others proceeding with outpatient evaluations. Defining the short-term prognosis and risk factors for stroke after TIA may provide guidance in determining which patients need rapid evaluation. OBJECTIVE: To determine the short-term risk of stroke and other adverse events after emergency department (ED) diagnosis of TIA. DESIGN AND SETTING: Cohort study conducted from March 1997 through February 1998 in 16 hospitals in a health maintenance organization in northern California. Patients A total of 1707 patients (mean age, 72 years) identified by ED physicians as having presented with TIA. MAIN OUTCOME MEASURES: Risk of stroke during the 90 days after index TIA; other events, including death, recurrent TIA, and hospitalization for cardiovascular events. RESULTS: During the 90 days after index TIA, 180 patients (10.5%) returned to the ED with a stroke, 91 of which occurred in the first 2 days. Five factors were independently associated with stroke: age greater than 60 years (odds ratio [OR], 1.8; 95% confidence interval [CI], 1.1-2.7; P=.01), diabetes mellitus (OR, 2.0; 95% CI, 1.4-2.9; P<.001), symptom duration longer than 10 minutes (OR, 2.3; 95% CI, 1.3-4.2; P=.005), weakness (OR, 1.9; 95% CI, 1.4-2.6; P<.001), and speech impairment (OR, 1.5; 95% CI, 1.1-2.1; P=.01). Stroke or other adverse events occurred in 428 patients (25.1%) in the 90 days after the TIA and included 44 hospitalizations for cardiovascular events (2.6%), 45 deaths (2.6%), and 216 recurrent TIAs (12.7%). CONCLUSIONS: Our results indicate that the short-term risk of stroke and other adverse events among patients who present to an ED with a TIA is substantial. Characteristics of the patient and the TIA may be useful for identifying patients who may benefit from expeditious evaluation and treatment.

Question: Prognosis

Study Type: Cohort

Learner Level: Beginner / Intermediate

Notes: Landmark study that changed thinking about otherwise 'uncomplicated' TIA's. Sound methodology. Good discussion points: An extra 'value added'-the authors assessed predictors of short term stroke and proposed a scoring system for the purpose of identifying a low-risk subgroup. This is the basis of a good entry level discussion of the difference between a validated prediction rule and the qualitative information embodied in identification of prognostic factors.

Larson, E. B., et al. (2004). "Survival after initial diagnosis of Alzheimer disease." *Ann Intern Med* 140(7): 501-509.

BACKGROUND: Alzheimer disease is an increasingly common condition in older people. Knowledge of life expectancy after the diagnosis of Alzheimer disease and of associations of patient characteristics with survival may help planning for future care. OBJECTIVE: To investigate the course of Alzheimer disease after initial diagnosis and examine associations hypothesized to correlate with survival among community-dwelling patients with Alzheimer disease. DESIGN: Prospective observational study. SETTING: An Alzheimer disease patient registry from a base population of 23 000 persons age 60 years and older in the Group Health Cooperative, Seattle, Washington. PATIENTS: 521 newly recognized persons with Alzheimer disease enrolled from 1987 to 1996 in an Alzheimer disease patient registry. MEASUREMENTS: Baseline measurements included patient demographic features, Mini-Mental State Examination score, Blessed Dementia Rating Scale score, duration since reported onset of symptoms, associated symptoms, comorbid conditions, and selected signs. Survival was the outcome of interest. RESULTS: The median survival from initial diagnosis was 4.2 years for men and 5.7 years for women with Alzheimer disease. Men had poorer survival across all age groups compared with females. Survival was decreased in all age groups compared with the life expectancy of the U.S. population. Predictors of mortality based on proportional hazards models included a baseline Mini-Mental State Examination score of 17 or less, baseline Blessed Dementia Rating Scale score of 5.0 or greater, presence of frontal lobe release signs, presence of extrapyramidal signs, gait disturbance, history of falls, congestive heart failure, ischemic heart disease, and diabetes at baseline. LIMITATIONS: The base population, although typical of the surrounding Seattle community, may not be representative of other, more diverse populations. CONCLUSIONS: In this sample of community-dwelling elderly persons who received a diagnosis of Alzheimer disease, survival duration was shorter than predicted on the basis of U.S. population data, especially for persons with onset at relatively younger ages. Features significantly associated with reduced survival at diagnosis were increased severity of cognitive impairment, decreased functional level, history of falls, physical examination findings of frontal release signs, and abnormal gait. The variables most strongly associated with survival were measures of disease severity at the time of diagnosis. These results should be useful to patients and families experiencing Alzheimer disease, other caregivers, clinicians, and policymakers when planning for future care needs.

Question: Prognosis

Study Type: Prospective Cohort

Learner Level: Beginner / Intermediate

Notes: Important article that provides key information on survival in Alzheimer's disease. Good Discussion Points: Nicely illustrates the challenge of creating an inception cohort in a disease like dementia. Provides clear examples of survival analysis and Kaplan-Meier curves. Generates great discussion about how to apply (and not apply) this data to individual patients.

Lim, W. S., et al. (2003). "Defining community acquired pneumonia severity on presentation to hospital: an international derivation and validation study." *Thorax* 58(5): 377-382.

BACKGROUND: In the assessment of severity in community acquired pneumonia (CAP), the modified British Thoracic Society (mBTS) rule identifies patients with severe pneumonia but not patients who might be suitable for home management. A multicentre study was conducted to derive and validate a practical severity assessment model for stratifying adults hospitalised with CAP into different management groups. **METHODS:** Data from three prospective studies of CAP conducted in the UK, New Zealand, and the Netherlands were combined. A derivation cohort comprising 80% of the data was used to develop the model. Prognostic variables were identified using multiple logistic regression with 30 day mortality as the outcome measure. The final model was tested against the validation cohort. **RESULTS:** 1068 patients were studied (mean age 64 years, 51.5% male, 30 day mortality 9%). Age \geq 65 years (OR 3.5, 95% CI 1.6 to 8.0) and albumin $<$ 30 g/dl (OR 4.7, 95% CI 2.5 to 8.7) were independently associated with mortality over and above the mBTS rule (OR 5.2, 95% CI 2.7 to 10). A six point score, one point for each of Confusion, Urea $>$ 7 mmol/l, Respiratory rate $>$ 30/min, low systolic ($<$ 90 mm Hg) or diastolic ($<$ 60 mm Hg) Blood pressure), age \geq 65 years (CURB-65 score) based on information available at initial hospital assessment, enabled patients to be stratified according to increasing risk of mortality: score 0, 0.7%; score 1, 3.2%; score 2, 3%; score 3, 17%; score 4, 41.5% and score 5, 57%. The validation cohort confirmed a similar pattern. **CONCLUSIONS:** A simple six point score based on confusion, urea, respiratory rate, blood pressure, and age can be used to stratify patients with CAP into different management groups.

Question: Prognosis

Study Type: Clinical Prediction Rule

Learner Level: Intermediate to Advanced

Notes: CURB-65 is an alternative to PORT scores for predicting severity of illness and to stratify groups by mortality risk. This is a well done, clearly reported study that enhances the original work done by Fine et al (Prediction rule to identify low-risk patients with community acquired pneumonia. *NEJM*: 1997:336:243-50).

Mukamal, K. J., et al. (2001). "Prior alcohol consumption and mortality following acute myocardial infarction." *JAMA* 285(15): 1965-1970.

CONTEXT: Studies have found that individuals who consume 1 alcoholic drink every 1 to 2 days have a lower risk of a first acute myocardial infarction (AMI) than abstainers or heavy drinkers, but the effect of prior drinking on mortality after AMI is uncertain. **OBJECTIVE:** To determine the effect of prior alcohol consumption on long-term mortality among early survivors of AMI. **DESIGN AND SETTING:** Prospective inception cohort study conducted at 45 US community and tertiary care hospitals between August 1989 and September 1994, with a median follow-up of 3.8 years. **PATIENTS:** A total of 1913 adults hospitalized with AMI between 1989 and 1994. **MAIN OUTCOME MEASURE:** All-cause mortality, compared by self-reported average weekly consumption of beer, wine, and liquor during the year prior to AMI. **RESULTS:** Of the 1913 patients, 896 (47%) abstained from alcohol, 696 (36%) consumed less than 7 alcoholic drinks/wk, and 321 (17%) consumed 7 or more alcoholic drinks/wk. Compared with abstainers, patients who consumed less than 7 drinks/wk had a lower all-cause mortality rate (3.4 vs 6.3 deaths per 100 person-years; hazard ratio [HR], 0.55; 95% confidence interval [CI], 0.43-0.71) as did those who consumed 7 or more drinks/wk (2.4 vs 6.3 deaths per 100 person-years; HR, 0.38; 95% CI, 0.25-0.55; $P <$.001 for trend). After adjusting for propensity to drink and other potential confounders, increasing alcohol consumption remained predictive of lower mortality for less than 7 drinks/wk, with an adjusted HR of 0.79 (95% CI, 0.60-1.03), and for 7 or more drinks/wk, with an adjusted HR of 0.68 (95% CI, 0.45-1.05; $P =$.01 for trend). The association was similar for total and cardiovascular mortality, among both men and women, and among different types of alcoholic beverages. **CONCLUSION:** Self-reported moderate alcohol consumption in the year prior to AMI is associated with reduced mortality following infarction.

Question: Prognosis

Study Type: Prospective Cohort

Learner Level: Beginner / Intermediate

Notes: These two articles (see also Abramson 2001) may be considered together (see sample case.) It is sometimes fun to compare and contrast two different articles about a similar prognosis topic. The populations studied through observational designs (in this case a cohort) are critical to how we understand and apply the information. Therefore, it is sometimes a good strategy to compare and contrast findings in studies done in different settings. In these two papers, there are many opportunities to discuss the impact of potential 'bias'; Try to consider why I chose these two papers- what about the two populations make for good comparison?? Note: either of these papers could also be looked at as a single paper.

Rao, S. V., et al. (2004). "Relationship of blood transfusion and clinical outcomes in patients with acute coronary syndromes." *JAMA* 292(13): 1555-1562.

CONTEXT: It is unclear if blood transfusion in anemic patients with acute coronary syndromes is associated with improved survival. OBJECTIVE: To determine the association between blood transfusion and mortality among patients with acute coronary syndromes who develop bleeding, anemia, or both during their hospital course. DESIGN, SETTING, AND PATIENTS: We analyzed 24,112 enrollees in 3 large international trials of patients with acute coronary syndromes (the GUSTO IIb, PURSUIT, and PARAGON B trials). Patients were grouped according to whether they received a blood transfusion during the hospitalization. The association between transfusion and outcome was assessed using Cox proportional hazards modeling that incorporated transfusion as a time-dependent covariate and the propensity to receive blood, and a landmark analysis. MAIN OUTCOME MEASURE: Thirty-day mortality. RESULTS: Of the patients included, 2401 (10.0%) underwent at least 1 blood transfusion during their hospitalization. Patients who underwent transfusion were older and had more comorbid illness at presentation and also had a significantly higher unadjusted rate of 30-day death (8.00% vs 3.08%; $P < .001$), myocardial infarction (MI) (25.16% vs 8.16%; $P < .001$), and death/MI (29.24% vs 10.02%; $P < .001$) compared with patients who did not undergo transfusion. Using Cox proportional hazards modeling that incorporated transfusion as a time-dependent covariate, transfusion was associated with an increased hazard for 30-day death (adjusted hazard ratio [HR], 3.94; 95% confidence interval [CI], 3.26-4.75) and 30-day death/MI (HR, 2.92; 95% CI, 2.55-3.35). In the landmark analysis that included procedures and bleeding events, transfusion was associated with a trend toward increased mortality. The predicted probability of 30-day death was higher with transfusion at nadir hematocrit values above 25%. CONCLUSIONS: Blood transfusion in the setting of acute coronary syndromes is associated with higher mortality, and this relationship persists after adjustment for other predictive factors and timing of events. Given the limitations of post hoc analysis of clinical trials data, a randomized trial of transfusion strategies is warranted to resolve the disparity in results between our study and other observational studies. We suggest caution regarding the routine use of blood transfusion to maintain arbitrary hematocrit levels in stable patients with ischemic heart disease.

Question: Prognosis

Study Type: Cohort

Learner Level: Intermediate

Notes: Clinical data from multiple prospective randomized trials were pooled to create a cohort to study the relationship between blood transfusion in anemic patients with acute coronary syndromes and survival. A strength of this as a teaching paper is that one can discuss the fact that RCTs can provide much additional data in addition to the intervention initially studied. In addition, the results are interesting in that they suggest increased mortality associated with transfusion in some patients. This is counter to the common practice of transfusing to "a number threshold" rather than treating the patient. Good paper to combine with Hebert et al NEJM 1999;340:409-417

Salazar-Martinez, E., et al. (2004). "Coffee consumption and risk for type 2 diabetes mellitus." *Ann Intern Med* 140(1): 1-8.

BACKGROUND: In small, short-term studies, acute administration of caffeine decreases insulin sensitivity and impairs glucose tolerance. OBJECTIVE: To examine the long-term relationship between consumption of coffee and other caffeinated beverages and incidence of type 2 diabetes mellitus. DESIGN: Prospective cohort study. SETTING: The Nurses' Health Study and Health Professionals' Follow-up Study. PARTICIPANTS: The authors followed 41 934 men from 1986 to 1998 and 84 276 women from 1980 to 1998. These participants did not have diabetes, cancer, or cardiovascular disease at baseline. MEASUREMENTS: Coffee consumption was assessed every 2 to 4 years through validated questionnaires. RESULTS: The authors documented 1333 new cases of type 2 diabetes in men and 4085 new cases in women. The authors found an inverse association between coffee intake and type 2 diabetes after adjustment for age, body mass index, and other risk factors. The multivariate relative risks for diabetes according to regular coffee consumption categories (0, <1, 1 to 3, 4 to 5, or > or =6 cups per day) in men were 1.00, 0.98, 0.93, 0.71, and 0.46 (95% CI, 0.26 to 0.82; $P = 0.007$ for trend), respectively. The corresponding multivariate relative risks in women were 1.00, 1.16, 0.99, 0.70, and 0.71 (CI, 0.56 to 0.89; $P < 0.001$ for trend), respectively. For decaffeinated coffee, the multivariate relative risks comparing persons who drank 4 cups or more per day with nondrinkers were 0.74 (CI, 0.48 to 1.12) for men and 0.85 (CI, 0.61 to 1.17) for women. Total caffeine intake from coffee and other sources was associated with a statistically significantly lower risk for diabetes in both men and women. CONCLUSIONS: These data suggest that long-term coffee consumption is associated with a statistically significantly lower risk for type 2 diabetes.

Question: Prognosis

Study Type: Prospective Cohort

Learner Level: Beginner / Advanced

Notes: This is a sample packet. See the entire teaching package that is in this section. Strong methods for a pair of well published prospective cohort trials with long term follow up (the nurses health study and the health professions follow-up study). Great topic, as coffee consumption is applicable to most folks, regardless of specialty. Good discussion points: Great examples of how the groups are not equal at the start of the trials and how adjusted analyses attempt to control for those differences. In addition, surprising findings in that the study designers had expected the opposite finding. 'Surprise' findings always make for fun teaching papers.

Scholes, D., et al. (2005). "Risk factors associated with acute pyelonephritis in healthy women." *Ann Intern Med* 142(1): 20-27.

BACKGROUND: Although most cases of acute pyelonephritis occur in otherwise healthy women, data on risk factors for this condition are lacking. OBJECTIVE: To evaluate infection characteristics, incidence, and risk factors associated with acute pyelonephritis in a sample of women. DESIGN: Population-based case-control study. SETTING: Group Health Cooperative, a prepaid health plan in Washington. PARTICIPANTS: 788 nonpregnant women, 18 to 49 years of age. Case-patients ($n = 242$) were women

with pyelonephritis who were identified from computerized databases. Controls were 546 similar-age women with no pyelonephritis diagnosis in the previous 5 years who were randomly selected from enrollment databases. Response rates for case-patients and controls were 73% and 64%, respectively. MEASUREMENTS: Characteristics of infection and potential risk factors for pyelonephritis, ascertained through computer-assisted telephone interview and computerized databases. RESULTS: 7% of case-patients were hospitalized. Escherichia coli was the infecting pathogen in 85% of cases. In multivariable models, factors associated with pyelonephritis risk were frequency of sexual intercourse in the previous 30 days (odds ratio, 5.6 [95% CI, 2.8 to 11.0] for > or =3 times per week), recent urinary tract infection (UTI) (odds ratio, 4.4 [CI, 2.8 to 7.1]), diabetes (odds ratio, 4.1 [CI, 1.6 to 10.9]), recent incontinence (odds ratio, 3.9 [CI, 2.6 to 5.9]), new sexual partner in the previous year (odds ratio, 2.2 [CI, 1.4 to 3.6]), recent spermicide use (odds ratio, 1.7 [CI, 1.1 to 2.8]), and UTI history in the participant's mother (odds ratio, 1.6 [CI, 1.1 to 2.5]). Risk factors for selected subgroups (patients < or = 30 years of age, patients > 30 years of age, patients with no UTI history, and inpatients) were also evaluated. LIMITATIONS: Potential recall bias, reliance on automated case definition criteria, and limited data on diabetes and incontinence variables. CONCLUSIONS: Few nonpregnant, community-dwelling women younger than 50 years of age with pyelonephritis are hospitalized. As with cystitis in reproductive-age women, sexual behaviors and patient and family history of UTI are associated with increased pyelonephritis risk. Diabetes and incontinence also seem to independently increase the risk for pyelonephritis.

Question: Prognosis

Study Type: Case-control

Learner Level: Beginner / Intermediate

Notes: Well-done case control study on a common condition. Transparent reporting of methods and results that show clear differences between the cases and controls makes this a very good case-control study to discuss, even for beginners. Good discussion points: As always with a case-control method, it is critical to get the learners to understand that a case control study begins with the identification of cases, who have already had the outcome. This is very clearly stated in this paper, as is the selection of controls. Another possible discussion point is the Managed Care setting (HMO in Seattle Washington). Could discuss a benefit of a managed care setting in which the databases are used to further understanding of risk factors for this very common problem.

Sears, M. R., et al. (2003). "A longitudinal, population-based, cohort study of childhood asthma followed to adulthood." N Engl J Med 349(15): 1414-1422.

BACKGROUND: The outcome of childhood asthma in adults has been described in high-risk cohorts, but few population-based studies have reported the risk factors for persistence and relapse. METHODS: We assessed children born from April 1972 through March 1973 in Dunedin, New Zealand, repeatedly from 9 to 26 years of age with questionnaires, pulmonary-function tests, bronchial-challenge testing, and allergy testing. RESULTS: By the age of 26 years, 51.4 percent of 613 study members with complete respiratory data had reported wheezing at more than one assessment. Eighty-nine study members (14.5 percent) had wheezing that persisted from childhood to 26 years of age, whereas 168 (27.4 percent) had remission, but 76 (12.4 percent) subsequently relapsed by the age of 26. Sensitization to house dust mites predicted the persistence of wheezing (odds ratio, 2.41; P=0.001) and relapse (odds ratio, 2.18; P=0.01), as did airway hyperresponsiveness (odds ratio for persistence, 3.00; P<0.001; odds ratio for relapse, 3.03; P<0.001). Female sex predicted the persistence of wheezing (odds ratio, 1.71; P=0.03), as did smoking at the age of 21 years (odds ratio, 1.84; P=0.01). The earlier the age at onset, the greater the risk of relapse (odds ratio, 0.89 per year of increase in the age at onset; P<0.001). Pulmonary function was consistently lower in those with persistent wheezing than in those without persistent wheezing. CONCLUSIONS: In an unselected birth cohort, more than one in four children had wheezing that persisted from childhood to adulthood or that relapsed after remission. The factors predicting persistence or relapse were sensitization to house dust mites, airway hyperresponsiveness, female sex, smoking, and early age at onset. These findings, together with persistently low lung function, suggest that outcomes in adult asthma may be determined primarily in early childhood.

Question: Prognosis

Study Type: Prospective Cohort

Learner Level: Beginner / Advanced

Notes: Well done prospective population based cohort. Good discussion points: Great for discussions of population selection for cohort studies and methodology for studies of prognosis. Tables and results are reasonably easy to read and understand for beginners.

Steinhauser, K. E., et al. (2000). "In search of a good death: observations of patients, families, and providers." Ann Intern Med 132(10): 825-832.

Despite a recent increase in the attention given to improving end-of-life care, our understanding of what constitutes a good death is surprisingly lacking. The purpose of this study was to gather descriptions of the components of a good death from patients, families, and providers through focus group discussions and in-depth interviews. Seventy-five participants-including physicians, nurses, social workers, chaplains, hospice volunteers, patients, and recently bereaved family members-were recruited from a university medical center, a Veterans Affairs medical center, and a community hospice. Participants identified six major components of a good death: pain and symptom management, clear decision making, preparation for death, completion, contributing to others, and affirmation of the whole person. The six themes are process-oriented attributes of a good death, and each has biomedical, psychological, social, and spiritual components. Physicians' discussions of a good death differed greatly from those of other groups. Physicians offered the most biomedical perspective, and patients, families, and other health care professionals defined a broad range of attributes integral

to the quality of dying. Although there is no "right" way to die, these six themes may be used as a framework for understanding what participants tend to value at the end of life. Biomedical care is critical, but it is only a point of departure toward total end-of-life care. For patients and families, psychosocial and spiritual issues are as important as physiologic concerns.

Question: Prognosis

Study Type: Qualitative / Focus groups and in-depth interviews

Learner Level: Beginner / Intermediate

Notes: This is a sample package with a few suggested teaching scenarios to help think about and/ or teach about qualitative methodology.

Tsugawa, Y., et al. (2017). "Physician age and outcomes in elderly patients in hospital in the US: observational study." *BMJ* 357: j1797.

Objectives To investigate whether outcomes of patients who were admitted to hospital differ between those treated by younger and older physicians. **Design** Observational study. **Setting** US acute care hospitals. **Participants** 20% random sample of Medicare fee-for-service beneficiaries aged ≥ 65 admitted to hospital with a medical condition in 2011-14 and treated by hospitalist physicians to whom they were assigned based on scheduled work shifts. To assess the generalizability of findings, analyses also included patients treated by general internists including both hospitalists and non-hospitalists. **Main outcome measures** 30 day mortality and readmissions and costs of care. **Results** 736 537 admissions managed by 18 854 hospitalist physicians (median age 41) were included. Patients' characteristics were similar across physician ages. After adjustment for characteristics of patients and physicians and hospital fixed effects (effectively comparing physicians within the same hospital), patients' adjusted 30 day mortality rates were 10.8% for physicians aged <40 (95% confidence interval 10.7% to 10.9%), 11.1% for physicians aged 40-49 (11.0% to 11.3%), 11.3% for physicians aged 50-59 (11.1% to 11.5%), and 12.1% for physicians aged ≥ 60 (11.6% to 12.5%). Among physicians with a high volume of patients, however, there was no association between physician age and patient mortality. Readmissions did not vary with physician age, while costs of care were slightly higher among older physicians. Similar patterns were observed among general internists and in several sensitivity analyses. **Conclusions** Within the same hospital, patients treated by older physicians had higher mortality than patients cared for by younger physicians, except those physicians treating high volumes of patients.

Question: Prognosis

Study Type: Observational study

Learner Level: Beginner

Notes: This is a good study for discussing observational data and associations. The conclusion (patients treated by older patients had higher mortality than patients cared for by younger patients) can be a great jumping point for discussion of confounders.

Venditti, M., et al. (2009). "Outcomes of patients hospitalized with community-acquired, health care-associated, and hospital-acquired pneumonia." *Ann Intern Med* 150(1): 19-26.

BACKGROUND: Traditionally, pneumonia has been classified as either community- or hospital-acquired. Although only limited data are available, health care-associated pneumonia has been recently proposed as a new category of respiratory infection. "Health care-associated pneumonia" refers to pneumonia in patients who have recently been hospitalized, had hemodialysis, or received intravenous chemotherapy or reside in a nursing home or long-term care facility. **OBJECTIVE:** To ascertain the epidemiology and outcome of community-acquired, health care-associated, and hospital-acquired pneumonia in adults hospitalized in internal medicine wards. **DESIGN:** Multicenter, prospective observational study. **SETTING:** 55 hospitals in Italy comprising 1941 beds. **PATIENTS:** 362 patients hospitalized with pneumonia during two 1-week surveillance periods. **MEASUREMENTS:** Cases of radiologically and clinically assessed pneumonia were classified as community-acquired, health care-associated, or hospital-acquired and rates were compared. **RESULTS:** Of the 362 patients, 61.6% had community-acquired pneumonia, 24.9% had health care-associated pneumonia, and 13.5% had hospital-acquired pneumonia. Patients with health care-associated pneumonia had higher mean Sequential Organ Failure Assessment scores than did those with community-acquired pneumonia (3.0 vs. 2.0), were more frequently malnourished (11.1% vs. 4.5%, and had more frequent bilateral (34.4% vs. 19.7%) and multilobar (27.8% vs. 21.5%) involvement on a chest radiograph. Patients with health care-associated pneumonia also had higher fatality rates (17.8% [CI, 10.6% to 24.9%] vs. 6.7% [CI, 2.9% to 10.5%]) and longer mean hospital stay (18.7 days [CI, 15.9 to 21.5 days] vs. 14.7 days [CI, 13.4 to 15.9 days]). Logistic regression analysis revealed that depression of consciousness (odds ratio [OR], 3.2 [CI, 1.06 to 9.8]), leukopenia (OR, 6.2 [CI, 1.01 to 37.6]), and receipt of empirical antibiotic therapy not recommended by international guidelines (OR, 6.4 [CI, 2.3 to 17.6]) were independently associated with increased intrahospital mortality. **Limitations:** The number of patients with health care-associated pneumonia was relatively small. Microbiological investigations were not always homogeneous. The study included only patients with pneumonia that required hospitalization; results may not apply to patients treated as outpatients. **CONCLUSION:** Health care-associated pneumonia should be considered a distinct subset of pneumonia associated with more severe disease, longer hospital stay, and higher mortality rates. Physicians should differentiate between patients with health care-associated pneumonia and those with community-acquired pneumonia and provide more appropriate initial antibiotic therapy.

Question: Prognosis

Study Type: Prospective Cohort

Learner Level: Intermediate

Notes: Good to teach prognosis

Vento, S., et al. (1998). "Fulminant hepatitis associated with hepatitis A virus superinfection in patients with chronic hepatitis C." N Engl J Med 338(5): 286-290.

BACKGROUND: Hepatitis A virus (HAV) infection rarely causes fulminant hepatic failure in people with no underlying liver disease. There are limited data on the course of this infection in patients with chronic hepatitis B and chronic hepatitis C. **METHODS:** We prospectively followed, from June 1990 to July 1997, 595 adults with biochemical and histologic evidence of chronic hepatitis B (163 patients) or chronic hepatitis C (432 patients) who were seronegative for HAV antibodies. All were tested every four months for serum IgM and IgG antibodies to HAV. **RESULTS:** Twenty-seven patients acquired HAV superinfection, 10 of whom had chronic hepatitis B and 17 of whom had chronic hepatitis C. One of the patients with chronic hepatitis B, who also had cirrhosis, had marked cholestasis (peak serum bilirubin level, 28 mg per deciliter [479 micromol per liter]); the other nine had uncomplicated courses of hepatitis A. Fulminant hepatic failure developed in seven of the patients with chronic hepatitis C, all but one of whom died. The other 10 patients with chronic hepatitis C had uncomplicated courses of hepatitis A. **CONCLUSIONS:** Although most patients with chronic hepatitis B who acquired HAV infection had an uncomplicated course, patients with chronic hepatitis C had a substantial risk of fulminant hepatitis and death associated with HAV superinfection. Our data suggest that patients with chronic hepatitis C should be vaccinated against hepatitis A.

Question: Prognosis

Study Type:

Learner Level: Beginner / Intermediate

Notes: This is a sample packet. See the entire teaching package that is in this section. This is a cohort of patients followed through a GI clinic. Good example of a prospective cohort in a referral center.

Weinstock, M. B., et al. (2015). "Risk for Clinically Relevant Adverse Cardiac Events in Patients With Chest Pain at Hospital Admission." JAMA Intern Med 175(7): 1207-1212.

IMPORTANCE: Patients with potentially ischemic chest pain are commonly admitted to the hospital or observed after a negative evaluation in the emergency department (ED) owing to concern about adverse events. Previous studies have looked at 30-day mortality, but no current large studies have examined the most important information regarding ED disposition: the short-term risk for a clinically relevant adverse cardiac event (including inpatient ST-segment elevation myocardial infarction, life-threatening arrhythmia, cardiac or respiratory arrest, or death). **OBJECTIVE:** To determine the incidence of clinically relevant adverse cardiac events in patients hospitalized for chest pain with 2 troponin-negative findings, nonconcerning initial ED vital signs, and nonischemic, interpretable electrocardiographic findings. **DESIGN, SETTING, AND PARTICIPANTS:** We conducted a blinded data review of 45,416 encounters obtained from a prospectively collected database enrolling adult patients admitted or observed with the following inclusion criteria: (1) primary presenting symptom of chest pain, chest tightness, chest burning, or chest pressure and (2) negative findings for serial biomarkers. Data were collected and analyzed from July 1, 2008, through June 30, 2013, from the EDs of 3 community teaching institutions with an aggregate census of more than 1 million visits. We analyzed data extracted by hypothesis-blinded abstractors. **MAIN OUTCOMES AND MEASURES:** The primary outcome was a composite of life-threatening arrhythmia, inpatient ST-segment elevation myocardial infarction, cardiac or respiratory arrest, or death during hospitalization. **RESULTS:** Of the 45,416 encounters, 11,230 met criteria for inclusion. Mean patient age was 58.0 years. Of the 11,230 encounters, 44.83% of patients arrived by ambulance and 55.00% of patients were women. Relevant history included hypertension in 46.00%, diabetes mellitus in 19.72%, and myocardial infarction in 13.16%. The primary end point occurred in 20 of the 11,230 patients (0.18% [95% CI, 0.11%-0.27%]). After excluding patients with abnormal vital signs, electrocardiographic ischemia, left bundle branch block, or a pacemaker rhythm, we identified a primary end point event in 4 of 7266 patients (0.06% [95% CI, 0.02%-0.14%]). Of these events, 2 were noncardiac and 2 were possibly iatrogenic. **CONCLUSIONS AND RELEVANCE:** In adult patients with chest pain admitted with 2 negative findings for serial biomarkers, nonconcerning vital signs, and nonischemic electrocardiographic findings, short-term clinically relevant adverse cardiac events were rare and commonly iatrogenic, suggesting that routine inpatient admission may not be a beneficial strategy for this group.

Question: Prognosis

Study Type: Retrospective

Learner Level: Advanced

Notes: Advanced Retrospective prognosis trial. Good for discussion on whether to admit/discharge patients with certain presenting signs being evaluated for chest pain.

Wisnivesky, J. P., et al. (2005). "Prospective validation of a prediction model for isolating inpatients with suspected pulmonary tuberculosis." Arch Intern Med 165(4): 453-457.

BACKGROUND: Current guidelines for the control of nosocomial transmission of tuberculosis (TB) recommend respiratory isolation for all patients with suspected TB. Application of these guidelines has resulted in many patients without TB being isolated on admission to the hospital, significantly increasing hospital costs. This study was conducted to prospectively validate a clinical decision rule to predict the need for respiratory isolation in inpatients with suspected TB. **METHODS:** A cohort of 516 individuals, who presented to 2 New York City hospitals between January 16, 2001, and September 29, 2002, and who were isolated on

admission for clinically suspected TB, were enrolled in the study. Face-to-face interviews were conducted to determine the presence of clinical variables associated with TB in the prediction model, including TB risk factors, clinical symptoms, and findings from physical examination and chest radiography. RESULTS: Of the 516 patients, 19 were found to have TB (prevalence, 3.7%; 95% confidence interval [CI], 2.2%-5.7%). The prediction rule had a sensitivity of 95% (95% CI, 74%-100%) and a specificity of 35% (95% CI, 31%-40%). Using a prevalence of TB of 3.7%, the positive predictive value was 9.6% and the negative predictive value was 99.7%. CONCLUSIONS: Among inpatients with suspected active pulmonary TB who are isolated on admission to the hospital, a prediction rule based on clinical and chest radiographic findings accurately identified patients at low risk for TB. Approximately one third of the unnecessary episodes of respiratory isolation could have been avoided had the prediction rule been applied. Future studies should assess the feasibility of implementing the rule in clinical practice.

Question: Prognosis

Study Type: Clinical Prediction Rule

Learner Level: Intermediate to Advanced

Notes: Well done validation study of a clinical prediction rule for identification of patients at low risk for TB in order to inform decisions regarding respiratory isolation. Good paper for introducing the concept of clinical prediction rules and their utility.

Syntheses - SR, MA, CPG

Annane, D., et al. (2004). "Corticosteroids for severe sepsis and septic shock: a systematic review and meta-analysis." *BMJ* 329(7464): 480.

OBJECTIVE: To assess the effects of corticosteroids on mortality in patients with severe sepsis and septic shock. **DATA SOURCES:** Randomised and quasi-randomised trials of corticosteroids versus placebo (or supportive treatment alone) retrieved from the Cochrane infectious diseases group's trials register, the Cochrane central register of controlled trials, Medline, Embase, and LILACS. **REVIEW METHOD:** Two pairs of reviewers agreed on eligibility of trials. One reviewer entered data on to the computer and four reviewers checked them. We obtained some missing data from authors of trials and assessed methodological quality of trials. **RESULTS:** 16/23 trials (n = 2063) were selected. Corticosteroids did not change 28 day mortality (15 trials, n = 2022; relative risk 0.92, 95% confidence interval 0.75 to 1.14) or hospital mortality (13 trials, n = 1418; 0.89, 0.71 to 1.11). There was significant heterogeneity. Subgroup analysis on long courses (> or = 5 days) with low dose (< or = 300 mg hydrocortisone or equivalent) corticosteroids showed no more heterogeneity. The relative risk for mortality was 0.80 at 28 days (five trials, n = 465; 0.67 to 0.95) and 0.83 at hospital discharge (five trials, n = 465, 0.71 to 0.97). Use of corticosteroids reduced mortality in intensive care units (four trials, n = 425, 0.83, 0.70 to 0.97), increased shock reversal at 7 days (four trials, n = 425; 1.60, 1.27 to 2.03) and 28 days (four trials, n = 425, 1.26, 1.04 to 1.52) without inducing side effects. **CONCLUSIONS:** For all trials, regardless of duration of treatment and dose, use of corticosteroids did not significantly affect mortality. With long courses of low doses of corticosteroids, however, mortality at 28 days and hospital mortality was reduced.

Question: Therapy

Study Type: Meta-analysis

Learner Level: Intermediate / Advanced

Notes: This is a well-done meta-analysis that can allow discussion of important concepts for intermediate to advanced learners who want to push their understanding of meta-analysis. *ACP Journal Club Summary* 2005; 142 (2): 30. Good discussion points: Can discuss both quasi-randomization (definition: pseudo random allocation method such as day of birth, order of participant enrollment in study, day of the week, medical record number etc.) and the introduction of possible selection bias. Can refer readers to tables 3 and 4 where specific design issues are identified. Can also discuss heterogeneity and the authors' attempts to explain it as multifactorial, including differences in methodologic strength and dose/ duration of intervention. Excellent review of how to think about subgroup analysis. Possible strengths (pre-stated hypothesis, statistically significant effect), and pitfalls (difference in comparisons between studies rather than within studies) of the interpretation of this subgroup analysis.

Bajwa, E. K., et al. (2005). "Interferon-gamma1b therapy in idiopathic pulmonary fibrosis: a metaanalysis." *Chest* 128(1): 203-206.

CONTEXT: Despite the investigation of multiple therapeutic options, idiopathic pulmonary fibrosis (IPF) remains a devastating, progressively fatal disease. Much interest has focused on the use of interferon (IFN)-gamma1b therapy, but the efficacy of this treatment has not been proven. **OBJECTIVE:** To determine whether IFN treatment reduces mortality in patients with IPF. **DESIGN:** A meta-analysis of randomized controlled trials evaluating the use of IFN-gamma1b as treatment for IPF. **MAIN OUTCOME MEASURE:** Mortality in patients treated with IFN-gamma1b was compared to mortality in patients treated with control therapies. **RESULTS:** A total of three studies involving 390 patients was included in the analysis. IFN-gamma1b therapy was associated with reduced mortality (hazard ratio [HR], 0.418; 95% confidence interval [CI], 0.253 to 0.690; p = 0.0003). A comparison of mortality at different time points revealed that IFN-gamma1b therapy was associated with significantly reduced mortality at 1 year (0.0861; 95% CI, 0.0244 to 0.1478; p = 0.0063), 18 months (0.1682; 95% CI, 0.1065 to 0.2299; p < 0.0001), 650 days (0.1939; 95% CI, 0.1386 to 0.2492; p < 0.0001), and 2 years (0.2652; 95% CI, 0.1652 to 0.3652; p < 0.0001). **CONCLUSION:** When the results of multiple studies are combined in a meta-analysis, IFN-gamma1b therapy is associated with reduced mortality.

Question: Therapy

Study Type: Meta-analysis

Learner Level: Intermediate

Notes: This can be taught as a package. First exercise goes through the user's guide for therapy and the first article. The second exercise goes through the user's guide for meta-analysis. It helps illustrate some of the points of the validity of the systematic review because the learner's have had a sneak peak at the underlying data. Teach with: Part 1: Raghu G et al. A Placebo-controlled trial of interferon gamma 1b in patients with idiopathic pulmonary fibrosis

Baker, W. L., et al. (2009). "Systematic review: comparative effectiveness of angiotensin-converting enzyme inhibitors or angiotensin II-receptor blockers for ischemic heart disease." *Ann Intern Med* 151(12): 861-871.

BACKGROUND: Patients with ischemic heart disease and preserved ventricular function experience considerable morbidity and mortality despite standard medical therapy. **PURPOSE:** To compare benefits and harms of using angiotensin-converting enzyme (ACE) inhibitors, angiotensin II-receptor blockers (ARBs), or combination therapy in adults with stable ischemic heart disease and preserved ventricular function. **DATA SOURCES:** MEDLINE, EMBASE, Cochrane Central Register of Controlled Trials, and Cochrane Database of Systematic Reviews (earliest date, July 2009) were searched without language restrictions. **STUDY SELECTION:** Two independent investigators screened citations for trials of at least 6 months' duration that compared ACE inhibitors, ARBs, or combination therapy with placebo or active control and reported any of several clinical outcomes. **DATA EXTRACTION:** Using standardized protocols, 2 independent investigators extracted information about study characteristics and rated the quality and strength of evidence. Disagreement was resolved by consensus. **DATA SYNTHESIS:** 41 studies met eligibility criteria. Moderate- to high-strength evidence (7 trials; 32 559 participants) showed that ACE inhibitors reduce the relative risk (RR) for total mortality (RR, 0.87 [95% CI, 0.81 to 0.94]) and nonfatal myocardial infarction (RR, 0.83 [CI, 0.73 to 0.94]) but increase the RR for syncope (RR, 1.24 [CI, 1.02 to 1.52]) and cough (RR, 1.67 [CI, 1.22 to 2.29]) compared with placebo. Low-strength evidence (1 trial; 5926 participants) suggested that ARBs reduce the RR for the composite end point of cardiovascular mortality, nonfatal myocardial infarction, or stroke (RR, 0.88 [CI, 0.77 to 1.00]) but not for the individual components. Moderate-strength evidence (1 trial; 25 620 participants) showed similar effects on total mortality (RR, 1.07 [CI, 0.98 to 1.16]) and myocardial infarction (RR, 1.08 [CI, 0.94 to 1.23]) but an increased risk for discontinuations because of hypotension ($P < 0.001$) and syncope ($P = 0.035$) with combination therapy compared with ACE inhibitors alone. **LIMITATIONS:** Many studies either did not assess or did not report harms in a systematic manner. Many studies did not adequately report benefits or harms by various patient subgroups. **CONCLUSION:** Adding an ACE inhibitor to standard medical therapy improves outcomes, including reduced risk for mortality and myocardial infarctions, in some patients with stable ischemic heart disease and preserved ventricular function. Less evidence supports a benefit of ARB therapy, and combination therapy seems no better than ACE inhibitor therapy alone and increases harms. **PRIMARY FUNDING SOURCE:** Agency for Healthcare Research and Quality.

Question: Therapy

Study Type: Systematic Review

Learner Level: Beginner

Notes: Well done meta-analysis; Nicely reported with transparent grading of evidence; Benefits nicely documented, but harms less so; Forest plots can be used to teach heterogeneity (Figure A. total mortality)

Bangalore, S., et al. (2008). "Perioperative beta blockers in patients having non-cardiac surgery: a meta-analysis." *Lancet* 372(9654): 1962-1976.

BACKGROUND: American College of Cardiology and American Heart Association (ACC/AHA) guidelines on perioperative assessment recommend perioperative beta blockers for non-cardiac surgery, although results of some clinical trials seem not to support this recommendation. We aimed to critically review the evidence to assess the use of perioperative beta blockers in patients having non-cardiac surgery. **METHODS:** We searched Pubmed and Embase for randomised controlled trials investigating the use of beta blockers in non-cardiac surgery. We extracted data for 30-day all-cause mortality, cardiovascular mortality, non-fatal myocardial infarction, non-fatal stroke, heart failure, and myocardial ischaemia, safety outcomes of perioperative bradycardia, hypotension, and bronchospasm. **FINDINGS:** 33 trials included 12 306 patients. beta blockers were not associated with any significant reduction in the risk of all-cause mortality, cardiovascular mortality, or heart failure, but were associated with a decrease (odds ratio [OR] 0.65, 95% CI 0.54-0.79) in non-fatal myocardial infarction (number needed to treat [NNT] 63) and decrease (OR 0.36, 0.26-0.50) in myocardial ischaemia (NNT 16) at the expense of an increase (OR 2.01, 1.27-3.68) in non-fatal strokes (number needed to harm [NNH] 293). The beneficial effects were driven mainly by trials with high risk of bias. For the safety outcomes, beta blockers were associated with a high risk of perioperative bradycardia requiring treatment (NNH 22), and perioperative hypotension requiring treatment (NNH 17). We recorded no increased risk of bronchospasm. **INTERPRETATION:** Evidence does not support the use of beta-blocker therapy for the prevention of perioperative clinical outcomes in patients having non-cardiac surgery. The ACC/AHA guidelines committee should soften their advocacy for this intervention until conclusive evidence is available.

Question: Therapy

Study Type: Meta-Analysis

Learner Level: .

Becker, G., et al. (2006). "Malignant ascites: systematic review and guideline for treatment." *Eur J Cancer* 42(5): 589-597.

A guideline on the management of symptomatic malignant ascites by abdominal paracentesis, diuretics and peritoneovenous shunting, based on a systematic review of the literature is presented. Thirty-two relevant studies were identified. None were randomized control trials, one was a non-randomized open controlled trial, five were cohort studies or prospective uncontrolled trials, 26 studies were non-analytic studies like case series. Although paracentesis, diuretics and shunting are commonly used procedures, the evidence is weak. Available data show good, although temporary effect of paracentesis on symptom relief. Fluid withdrawal speed and concurrent intravenous hydration is not sufficiently studied. Peritoneovenous shunts can control ascites in patients with malignant ascites, but have to be balanced by the potential risks of this procedure. The available data about diuretics in treatment of malignant ascites are controversial. The use of diuretics therefore should be considered in all patients, but has to be evaluated individually.

Question: Therapy

Study Type: Systematic Review

Learner Level: Beginner

Notes: Good to teach difference between a meta-analysis and a systematic review

Bent, S., et al. (1999). "Antibiotics in acute bronchitis: a meta-analysis." *Am J Med* 107(1): 62-67.

PURPOSE: Most patients with acute bronchitis who seek medical care are treated with antibiotics, although the effectiveness of this intervention is uncertain. We performed a meta-analysis of randomized, controlled trials to estimate the effectiveness of antibiotics in the treatment of acute bronchitis. **SUBJECTS AND METHODS:** English-language studies published January 1966 to April 1998 were retrieved using MEDLINE, bibliographies, and consultation with experts. Only randomized trials that enrolled otherwise healthy patients with a diagnosis of acute bronchitis, used an antibiotic in the treatment group and a placebo in the control group, and provided sufficient data to calculate an effect size were included. **RESULTS:** We identified eight randomized controlled trials that satisfied all inclusion criteria. These studies used one of three antibiotics (erythromycin, doxycycline, trimethoprim/sulfamethoxazole). The use of antibiotics decreased the duration of cough and sputum production by approximately one-half day (summary effect size 0.21; 95% CI, 0.05 to 0.36). For specific symptoms, there were nonsignificant trends favoring the use of antibiotics: a decrease of 0.4 days of purulent sputum (95% CI, -0.1 to 0.8), a decrease of 0.5 days of cough (95% CI, -0.1 to 1.1), and a decrease of 0.3 days lost from work (95% CI, -0.6 to 1.1). **CONCLUSION:** This meta-analysis suggests a small benefit from the use of the antibiotics erythromycin, doxycycline, or trimethoprim/sulfamethoxazole in the treatment of acute bronchitis in otherwise healthy patients. As this small benefit must be weighed against the risk of side effects and the societal cost of increasing antibiotic resistance, we believe that the use of antibiotics is not justified in these patients.

Question: Therapy

Study Type: Systematic Review

Learner Level: Intermediate

Notes: There are nice tables and figures that are classic for systematic review and meta-analysis. Good discussion points: This is a good paper for discussion of 'effect size,' (what you do when the papers in your study don't measure the same outcomes..) However, this may slightly confuse those who are not fully comfortable with meta-analysis.

Bischoff-Ferrari, H. A., et al. (2009). "Fall prevention with supplemental and active forms of vitamin D: a meta-analysis of randomised controlled trials." *BMJ* 339: b3692.

OBJECTIVE: To test the efficacy of supplemental vitamin D and active forms of vitamin D with or without calcium in preventing falls among older individuals. **DATA SOURCES:** We searched Medline, the Cochrane central register of controlled trials, BIOSIS, and Embase up to August 2008 for relevant articles. Further studies were identified by consulting clinical experts, bibliographies, and abstracts. We contacted authors for additional data when necessary. **Review methods** Only double blind randomised controlled trials of older individuals (mean age 65 years or older) receiving a defined oral dose of supplemental vitamin D (vitamin D(3) (cholecalciferol) or vitamin D(2) (ergocalciferol)) or an active form of vitamin D (1alpha-hydroxyvitamin D(3) (1alpha-hydroxycalciferol) or 1,25-dihydroxyvitamin D(3) (1,25-dihydroxycholecalciferol)) and with sufficiently specified fall assessment were considered for inclusion. **RESULTS:** Eight randomised controlled trials (n=2426) of supplemental vitamin D met our inclusion criteria. Heterogeneity among trials was observed for dose of vitamin D (700-1000 IU/day v 200-600 IU/day; P=0.02) and achieved 25-hydroxyvitamin D(3) concentration (25(OH)D concentration: <60 nmol/l v >or=60 nmol/l; P=0.005). High dose supplemental vitamin D reduced fall risk by 19% (pooled relative risk (RR) 0.81, 95% CI 0.71 to 0.92; n=1921 from seven trials), whereas achieved serum 25(OH)D concentrations of 60 nmol/l or more resulted in a 23% fall reduction (pooled RR 0.77, 95% CI 0.65 to 0.90). Falls were not notably reduced by low dose supplemental vitamin D (pooled RR 1.10, 95% CI 0.89 to 1.35; n=505 from two trials) or by achieved serum 25-hydroxyvitamin D concentrations of less than 60 nmol/l (pooled RR 1.35, 95% CI 0.98 to 1.84). Two randomised controlled trials (n=624) of active forms of vitamin D met our inclusion criteria. Active forms of vitamin D reduced fall risk by 22% (pooled RR 0.78, 95% CI 0.64 to 0.94). **CONCLUSIONS:** Supplemental vitamin D in a dose of 700-1000 IU a day reduced the risk of falling among older individuals by 19% and to a similar degree as active forms of vitamin D. Doses of supplemental vitamin D of less than 700 IU or serum 25-hydroxyvitamin D concentrations of less than 60 nmol/l may not reduce the risk of falling among older individuals.

Question: Therapy

Study Type: Meta-analysis

Learner Level: Intermediate

Notes: Nicely reported meta-analysis that can be used to illustrate issues of heterogeneity. When all studies were included in the combined outcome, the results showed significant heterogeneity. However, if the data are separated into high dose and low dose, the heterogeneity is significantly improved. There is an ACP Journal club summary for this article.

Briel, M., et al. (2018). "Corticosteroids in Patients Hospitalized With Community-Acquired Pneumonia: Systematic Review and Individual Patient Data Metaanalysis." *Clinical Infectious Diseases* 66(3): 346-354.

Background: Our aim was to evaluate the benefits and harms of adjunctive corticosteroids in adults hospitalized with community-acquired pneumonia (CAP) using individual patient data from randomized, placebo-controlled trials and to explore subgroup differences. **Methods:** We systematically searched Medline, Embase, Cochrane Central, and trial registers (all through July 2017). Data from 1506 individual patients in 6 trials were analyzed using uniform outcome definitions. We investigated prespecified effect modifiers using multivariable hierarchical regression, adjusting for pneumonia severity, age, and clustering effects. **Results:** Within 30 days of randomization, 37 of 748 patients (5.0%) assigned to corticosteroids and 45 of 758 patients (5.9%) assigned to placebo died (adjusted odds ratio [aOR], 0.75; 95% confidence interval [CI], .46 to 1.21; $P = .24$). Time to clinical stability and length of hospital stay were reduced by approximately 1 day with corticosteroids (-1.03 days; 95% CI, -1.62 to -.43; $P = .001$ and -1.15 days; 95% CI, -1.75 to -.55; $P < .001$, respectively). More patients with corticosteroids had hyperglycemia (160 [22.1%] vs 88 [12.0%]; aOR, 2.15; 95% CI, 1.60 to 2.90; $P < .001$) and CAP-related rehospitalization (33 [5.0%] vs 18 [2.7%]; aOR, 1.85; 95% CI, 1.03 to 3.32; $P = .04$). We did not find significant effect modification by CAP severity or degree of inflammation. **Conclusions:** Adjunct corticosteroids for patients hospitalized with CAP reduce time to clinical stability and length of hospital stay by approximately 1 day without a significant effect on overall mortality but with an increased risk for CAP-related rehospitalization and hyperglycemia.

Question: Therapy

Study Type: Systematic Review

Learner Level: Intermediate

Notes: Teach this paper in conjunction with the Cochrane Systematic Review by Stern et al. in order to demonstrate that systematic reviews done by different authorities may have conflicting results and that it is important to understand the methods the authors use to arrive at their conclusions.

Bruckert, E., et al. (2010). "Meta-analysis of the effect of nicotinic acid alone or in combination on cardiovascular events and atherosclerosis." *Atherosclerosis* 210(2): 353-361.

OBJECTIVE: High-density lipoprotein cholesterol (HDL-C) concentration is a strong predictor of cardiovascular events in both naive and statin-treated patients. Nicotinic acid is an attractive option for decreasing residual risk in statin-treated or statin-intolerant patients since it increases HDL-C by up to 20% and decreases low-density lipoprotein cholesterol and lipoprotein(a) plasma concentrations. **METHODS:** We performed a computerized PubMed literature search that focused on clinical trials evaluating niacin, alone or in combination with other lipid-lowering drugs, published between January 1966 and August 2008. **RESULTS:** Among 587 citations, 29 full articles were read and 14 were eligible for inclusion. Overall 11 randomized controlled trials enrolled 2682 patients in the active group and 3934 in the control group. In primary analysis, niacin significantly reduced major coronary events (relative odds reduction=25%, 95% CI 13, 35), stroke (26%, 95% CI=8, 41) and any cardiovascular events (27%, 95% CI=15, 37). Except for stroke, the pooled between-group difference remained significant in sensitivity analysis excluding the largest trial. In comparison with the non-niacin group, more patients in the niacin group had regression of coronary atherosclerosis (relative increase=92%, 95% CI=39, 67) whereas the rate of patients with progression decreased by 41%, 95% CI=25, 53. Similar effects of niacin were found on carotid intima thickness with a weighted mean difference in annual change of -17 microm/year (95% CI=-22, -12). **CONCLUSIONS:** Although the studies were conducted before statin therapy become standard care, and mostly in patients in secondary prevention, with various dosages of nicotinic acid 1-3g/day, this meta-analysis found positive effects of niacin alone or in combination on all cardiovascular events and on atherosclerosis evolution.

Question: Therapy

Study Type: Meta-analysis

Learner Level: intermediate

Notes: The title sounds impressive but there are all sorts of problems with it --- too broad a question, pooling with too much clinical & statistical heterogeneity, and weird use of OR. I would say this is a more difficult paper to use in groups who already understand the basics of what to look for.

Call, S. A., et al. (2005). "Does this patient have influenza?" *JAMA* 293(8): 987-997.

CONTEXT: Influenza vaccination lowers, but does not eliminate, the risk of influenza. Making a reliable, rapid clinical diagnosis is essential to appropriate patient management that may be especially important during shortages of antiviral agents caused by high demand. OBJECTIVES: To systematically review the precision and accuracy of symptoms and signs of influenza. A secondary objective was to review the operating characteristics of rapid diagnostic tests for influenza (results available in <30 min). DATA SOURCES: Structured search strategy using MEDLINE (January 1966-September 2004) and subsequent searches of bibliographies of retrieved articles to identify articles describing primary studies dealing with the diagnosis of influenza based on clinical signs and symptoms. The MEDLINE search used the Medical Subject Headings EXP influenza or EXP influenza A virus or EXP influenza A virus human or EXP influenza B virus and the Medical Subject Headings or terms EXP sensitivity and specificity or EXP medical history taking or EXP physical examination or EXP reproducibility of results or EXP observer variation or symptoms.mp or clinical signs.mp or sensitivity.mp or specificity.mp. STUDY SELECTION: Of 915 identified articles on clinical assessment of influenza-related illness, 17 contained data on the operating characteristics of symptoms and signs using an independent criterion standard. Of these, 11 were eliminated based on 4 inclusion criteria and availability of nonduplicative primary data. DATA EXTRACTION: Two authors independently reviewed and abstracted data for estimating the likelihood ratios (LRs) of clinical diagnostic findings. Differences were resolved by discussion and consensus. DATA SYNTHESIS: No symptom or sign had a summary LR greater than 2 in studies that enrolled patients without regard to age. For decreasing the likelihood of influenza, the absence of fever (LR, 0.40; 95% confidence interval [CI], 0.25-0.66), cough (LR, 0.42; 95% CI, 0.31-0.57), or nasal congestion (LR, 0.49; 95% CI, 0.42-0.59) were the only findings that had summary LRs less than 0.5. In studies limited to patients aged 60 years or older, the combination of fever, cough, and acute onset (LR, 5.4; 95% CI, 3.8-7.7), fever and cough (LR, 5.0; 95% CI, 3.5-6.9), fever alone (LR, 3.8; 95% CI, 2.8-5.0), malaise (LR, 2.6; 95% CI, 2.2-3.1), and chills (LR, 2.6; 95% CI, 2.0-3.2) increased the likelihood of influenza to the greatest degree. The presence of sneezing among older patients made influenza less likely (LR, 0.47; 95% CI, 0.24-0.92). CONCLUSIONS: Clinical findings identify patients with influenza-like illness but are not particularly useful for confirming or excluding the diagnosis of influenza. Clinicians should use timely epidemiologic data to ascertain if influenza is circulating in their communities, then either treat patients with influenza-like illness empirically or obtain a rapid influenza test to assist with management decisions.

Question: Diagnosis

Study Type: Meta-analysis

Learner Level: Intermediate / Advanced

Notes: This meta-analysis is part of the Rational Clinical Exam Series, which reports information on the diagnostic test characteristics of history and physical exam items as well as a limited number of associated diagnostic tests. This article on influenza is timely in that the distribution of flu vaccine was altered by the failure of one European manufacturer to provide expected doses of vaccine to the United States. Thus, this paper was published in a setting of heightened public awareness of risk of influenza. Good discussion points: As with all Rational Clinical Exam articles, this is an evidence summary of diagnostic tests- thus one can discuss both systematic review methodology AND diagnosis, specifically Likelihood ratio. Excellent paragraph under statistical methods (Page 990) that defines likelihood ratio as well as diagnostic odds ratio. Stumbling block may be the large number of items listed in Tables 2 and 3 (p 992 and 993), however a clear difference in data can be noted in patients older than 60 years compared to all comers.

Carlberg, B., et al. (2004). "Atenolol in hypertension: is it a wise choice?" *Lancet* 364(9446): 1684-1689.

BACKGROUND: Atenolol is one of the most widely used beta blockers clinically, and has often been used as a reference drug in randomised controlled trials of hypertension. However, questions have been raised about atenolol as the best reference drug for comparisons with other antihypertensives. Thus, our aim was to systematically review the effect of atenolol on cardiovascular morbidity and mortality in hypertensive patients. METHODS: Reports were identified through searches of The Cochrane Library, MEDLINE, relevant textbooks, and by personal communication with established researchers in hypertension. Randomised controlled trials that assessed the effect of atenolol on cardiovascular morbidity or mortality in patients with primary hypertension were included. FINDINGS: We identified four studies that compared atenolol with placebo or no treatment, and five that compared atenolol with other antihypertensive drugs. Despite major differences in blood pressure lowering, there were no outcome differences between atenolol and placebo in the four studies, comprising 6825 patients, who were followed up for a mean of 4.6 years on all-cause mortality (relative risk 1.01 [95% CI 0.89-1.15]), cardiovascular mortality (0.99 [0.83-1.18]), or myocardial infarction (0.99 [0.83-1.19]). The risk of stroke, however, tended to be lower in the atenolol than in the placebo group (0.85 [0.72-1.01]). When atenolol was compared with other antihypertensives, there were no major differences in blood pressure lowering between the treatment arms. Our meta-analysis showed a significantly higher mortality (1.13 [1.02-1.25]) with atenolol treatment than with other active treatment, in the five studies comprising 17671 patients who were followed up for a mean of 4.6 years. Moreover, cardiovascular mortality also tended to be higher with atenolol treatment than with other antihypertensive treatment. Stroke was also more frequent with atenolol treatment. INTERPRETATION: Our results cast doubts on atenolol as a suitable drug for hypertensive patients. Moreover, they challenge the use of atenolol as a reference drug in outcome trials in hypertension.

Question: Therapy

Study Type: Meta-analysis

Learner Level: Beginner

Notes: This is a nice straightforward meta-analysis that shows how a commonly used drug may not end up providing as much clinical benefit as originally thought. Can also be used to discuss how a surrogate end-point (blood pressure) does not always reflect the true impact on clinical events.

Cipriani, A., et al. (2018). "Comparative efficacy and acceptability of 21 antidepressant drugs for the acute treatment of adults with major depressive disorder: a systematic review and network meta-analysis." *Lancet* 391(10128): 1357-1366.

BACKGROUND: Major depressive disorder is one of the most common, burdensome, and costly psychiatric disorders worldwide in adults. Pharmacological and non-pharmacological treatments are available; however, because of inadequate resources, antidepressants are used more frequently than psychological interventions. Prescription of these agents should be informed by the best available evidence. Therefore, we aimed to update and expand our previous work to compare and rank antidepressants for the acute treatment of adults with unipolar major depressive disorder. **METHODS:** We did a systematic review and network meta-analysis. We searched Cochrane Central Register of Controlled Trials, CINAHL, Embase, LILACS database, MEDLINE, MEDLINE In-Process, PsycINFO, the websites of regulatory agencies, and international registers for published and unpublished, double-blind, randomised controlled trials from their inception to Jan 8, 2016. We included placebo-controlled and head-to-head trials of 21 antidepressants used for the acute treatment of adults (≥ 18 years old and of both sexes) with major depressive disorder diagnosed according to standard operationalised criteria. We excluded quasi-randomised trials and trials that were incomplete or included 20% or more of participants with bipolar disorder, psychotic depression, or treatment-resistant depression; or patients with a serious concomitant medical illness. We extracted data following a predefined hierarchy. In network meta-analysis, we used group-level data. We assessed the studies' risk of bias in accordance to the Cochrane Handbook for Systematic Reviews of Interventions, and certainty of evidence using the Grading of Recommendations Assessment, Development and Evaluation framework. Primary outcomes were efficacy (response rate) and acceptability (treatment discontinuations due to any cause). We estimated summary odds ratios (ORs) using pairwise and network meta-analysis with random effects. This study is registered with PROSPERO, number CRD42012002291. **FINDINGS:** We identified 28 552 citations and of these included 522 trials comprising 116 477 participants. In terms of efficacy, all antidepressants were more effective than placebo, with ORs ranging between 2.13 (95% credible interval [CrI] 1.89-2.41) for amitriptyline and 1.37 (1.16-1.63) for reboxetine. For acceptability, only agomelatine (OR 0.84, 95% CrI 0.72-0.97) and fluoxetine (0.88, 0.80-0.96) were associated with fewer dropouts than placebo, whereas clomipramine was worse than placebo (1.30, 1.01-1.68). When all trials were considered, differences in ORs between antidepressants ranged from 1.15 to 1.55 for efficacy and from 0.64 to 0.83 for acceptability, with wide CrIs on most of the comparative analyses. In head-to-head studies, agomelatine, amitriptyline, escitalopram, mirtazapine, paroxetine, venlafaxine, and vortioxetine were more effective than other antidepressants (range of ORs 1.19-1.96), whereas fluoxetine, fluvoxamine, reboxetine, and trazodone were the least efficacious drugs (0.51-0.84). For acceptability, agomelatine, citalopram, escitalopram, fluoxetine, sertraline, and vortioxetine were more tolerable than other antidepressants (range of ORs 0.43-0.77), whereas amitriptyline, clomipramine, duloxetine, fluvoxamine, reboxetine, trazodone, and venlafaxine had the highest dropout rates (1.30-2.32). 46 (9%) of 522 trials were rated as high risk of bias, 380 (73%) trials as moderate, and 96 (18%) as low; and the certainty of evidence was moderate to very low. **INTERPRETATION:** All antidepressants were more efficacious than placebo in adults with major depressive disorder. Smaller differences between active drugs were found when placebo-controlled trials were included in the analysis, whereas there was more variability in efficacy and acceptability in head-to-head trials. These results should serve evidence-based practice and inform patients, physicians, guideline developers, and policy makers on the relative merits of the different antidepressants. **FUNDING:** National Institute for Health Research Oxford Health Biomedical Research Centre and the Japan Society for the Promotion of Science.

Question: Therapy

Study Type: Network Meta-Analysis

Learner Level: Advanced

Notes:

Colli, A., et al. (2006). "Accuracy of ultrasonography, spiral CT, magnetic resonance, and alpha-fetoprotein in diagnosing hepatocellular carcinoma: a systematic review." *Am J Gastroenterol* 101(3): 513-523.

BACKGROUND AND AIM: In patients with chronic liver disease, the accuracy of ultrasound scan (US), spiral computed tomography (CT), magnetic resonance imaging (MRI), and alpha-fetoprotein (AFP) in diagnosing hepatocellular carcinoma (HCC) has never been systematically assessed, and present systematic review was aimed at this issue. **METHODS:** Pertinent cross-sectional studies having as a reference standard pathological examinations of the explanted liver or resected segment(s), biopsies of focal lesion(s), and/or a period of follow-up, were identified using MEDLINE, EMBASE, Cochrane Library, and CancerLit. Pooled sensitivity, specificity, and likelihood ratios (LR) were calculated using the random effect model. Summary receiver operating characteristic (SROC) curve and predefined subgroup analyses were made when indicated. **RESULTS:** The pooled estimates of the 14 US studies were 60% (95% CI 44-76) for sensitivity, 97% (95% CI 95-98) for specificity, 18 (95% CI 8-37) for LR+, and 0.5 (95% CI 0.4-0.6) for LR-; for the 10 CT studies sensitivity was 68% (95% CI 55-80), specificity 93% (95% CI 89-96), LR+ 6 (95% CI 3-12), and LR- 0.4 (95% CI 0.3-0.6); for the nine MRI studies sensitivity was 81% (95% CI 70-91), specificity 85% (95% CI 77-93), LR+ 3.9 (95% CI 2-7), and LR- 0.3 (95% CI 0.2-0.5). The sensitivity and specificity of AFP varied widely, and this could not be entirely attributed to the threshold effect of the different cutoff levels used. **CONCLUSIONS:** US is highly specific but insufficiently sensitive to detect HCC in many cirrhotics or to support an effective surveillance program. The operative characteristics of CT are comparable, whereas MRI is more sensitive. High-quality prospective studies are needed to define the actual diagnostic role of AFP.

Question: Diagnosis

Study Type: Systematic Review

Learner Level: Advanced

Notes: Diagnostic tests for hepatocellular carcinoma; Lots of data – you can do calculation of LR from this paper but it will require very much set up first

However, it is a good paper for comparing likelihood ratios, discussing positive and negative LRs; Also can discuss the different cut-offs used for continuous data; Has an accompanying ACP-JC would could easily be used to simplify an exercise on this topic

Cramer, H., et al. (2013). "A systematic review and meta-analysis of yoga for low back pain." *Clinical Journal of Pain* 29(5): 450-460.

OBJECTIVES: To systematically review and meta-analyze the effectiveness of yoga for low back pain. **METHODS:** MEDLINE, the Cochrane Library, EMBASE, CAMBASE, and PsycINFO, were screened through January 2012. Randomized controlled trials comparing yoga to control conditions in patients with low back pain were included. Two authors independently assessed risk of bias using the risk of bias tool recommended by the Cochrane Back Review Group. Main outcome measures were pain, back-specific disability, generic disability, health-related quality of life, and global improvement. For each outcome, standardized mean differences (SMD) and 95% confidence intervals (CI) were calculated. **RESULTS:** Ten randomized controlled trials with a total of 967 chronic low back pain patients were included. Eight studies had low risk of bias. There was strong evidence for short-term effects on pain (SMD=-0.48; 95% CI, -0.65 to -0.31; P<0.01), back-specific disability (SMD=-0.59; 95% CI, -0.87 to -0.30; P<0.01), and global improvement (risk ratio=3.27; 95% CI, 1.89-5.66; P<0.01). There was strong evidence for a long-term effect on pain (SMD=-0.33; 95% CI, -0.59 to -0.07; P=0.01) and moderate evidence for a long-term effect on back-specific disability (SMD=-0.35; 95% CI, -0.55 to -0.15; P<0.01). There was no evidence for either short-term or long-term effects on health-related quality of life. Yoga was not associated with serious adverse events. **DISCUSSION:** This systematic review found strong evidence for short-term effectiveness and moderate evidence for long-term effectiveness of yoga for chronic low back pain in the most important patient-centered outcomes. Yoga can be recommended as an additional therapy to chronic low back pain patients.

Question: Therapy

Study Type: Meta-Analysis

Learner Level: Beginner

Notes: Beginner to intermediate level, simple, well-done meta-analysis of a common question, easy to teach I2, instructive Forest plots.

De Berardis, G., et al. (2009). "Aspirin for primary prevention of cardiovascular events in people with diabetes: meta-analysis of randomised controlled trials." *BMJ* 339: b4531.

OBJECTIVE: To evaluate the benefits and harms of low dose aspirin in people with diabetes and no cardiovascular disease. **DESIGN:** Meta-analysis of randomised controlled trials. **DATA SOURCES:** Medline (1966-November 2008), the Cochrane central register of controlled trials (Cochrane Library 2008;issue 4), and reference lists of retrieved articles. Review methods Randomised trials of aspirin compared with placebo or no aspirin in people with diabetes and no pre-existing cardiovascular disease were eligible for inclusion. Data on major cardiovascular events (death from cardiovascular causes, non-fatal myocardial infarction, non-fatal stroke, and all cause mortality) were extracted and pooled with a random effect model. Results are reported as relative risks with 95% confidence intervals. **RESULTS:** Of 157 studies in the literature searches, six were eligible (10 117 participants). When aspirin was compared with placebo there was no statistically significant reduction in the risk of major cardiovascular events (five studies, 9584 participants; relative risk 0.90, 95% confidence interval 0.81 to 1.00), cardiovascular mortality (four studies, n=8557, 0.94; 0.72 to 1.23), or all cause mortality (four studies, n=8557; 0.93, 0.82 to 1.05). Significant heterogeneity was found in the analysis for myocardial infarction (I(2)=62.2%; P=0.02) and stroke (I(2)=52.5%; P=0.08). Aspirin significantly reduced the risk of myocardial infarction in men (0.57, 0.34 to 0.94) but not in women (1.08, 0.71 to 1.65; P for interaction=0.056). Evidence relating to harms was inconsistent. **CONCLUSIONS:** A clear benefit of aspirin in the primary prevention of major cardiovascular events in people with diabetes remains unproved. Sex may be an important effect modifier. Toxicity is to be explored further.

Question: Therapy

Study Type: Meta-analysis

Learner Level: Beginner - Intermediate

Notes: Beginner / Intermediate meta-analysis of aspirin for primary prevention in diabetics; Large number of patients, nice forest plots; Very good for teaching about heterogeneity and I-square

de Ferranti, S. D., et al. (1998). "Are amoxicillin and folate inhibitors as effective as other antibiotics for acute sinusitis? A meta-analysis." *BMJ* 317(7159): 632-637.

OBJECTIVES: To examine whether antibiotics are indicated in treating uncomplicated acute sinusitis and, if so, whether newer and more expensive antibiotics with broad spectra of antimicrobial activity are more effective than amoxicillin or folate inhibitors. **DESIGN:** Meta-analysis of randomised trials. **SETTING:** Outpatient clinics. **SUBJECTS:** 2717 patients with acute sinusitis or acute exacerbation of chronic sinusitis from 27 trials. **INTERVENTIONS:** Any antibiotic versus placebo; amoxicillin or folate inhibitors versus newer, more expensive antibiotics. **MAIN OUTCOME MEASUREMENTS:** Clinical failures and cures. **RESULTS:** Compared with placebo, antibiotics decreased the incidence of clinical failures by half (risk ratio 0.54 (95% confidence interval 0.37 to 0.79)). Risk of clinical failure among 1553 randomised patients was not meaningfully decreased with more expensive antibiotics as compared with amoxicillin (risk ratio 0.86 (0.62 to 1.19); risk difference 0.9 fewer failures per 100 patients (1.4 more failures to 3.1 fewer failures per 100 patients)). The results were similar for other antibiotics versus folate inhibitors (risk ratio 1.01 (0.52 to 1.97)), but data were sparse (n=410) and of low quality. **CONCLUSIONS:** Amoxicillin and folate inhibitors are essentially as effective as more expensive

antibiotics for the initial treatment of uncomplicated acute sinusitis. Small differences in efficacy may exist, but are unlikely to be clinically important.

Question: Therapy

Study Type: Systematic Review

Learner Level: Beginner / Intermediate

Notes: There are nice tables and figures that are classic for systematic review and meta-analysis. Good discussion points: Good discussion of both a positive finding (antibiotics vs. placebo) and also negative finding (cheap vs. expensive antibiotics). Also good to discuss cost implications.

Dentali, F., et al. (2007). "Meta-analysis: anticoagulant prophylaxis to prevent symptomatic venous thromboembolism in hospitalized medical patients." *Ann Intern Med* 146(4): 278-288.

BACKGROUND: Underutilization of anticoagulant prophylaxis may be due to lack of evidence that prophylaxis prevents clinically important outcomes in hospitalized medical patients at risk for venous thromboembolism. **PURPOSE:** To assess the effects of anticoagulant prophylaxis in reducing clinically important outcomes in hospitalized medical patients. **DATA SOURCES:** MEDLINE, EMBASE, and Cochrane databases were searched to September 2006 without language restrictions. **STUDY SELECTION:** Randomized trials comparing anticoagulant prophylaxis with no treatment in hospitalized medical patients. **DATA EXTRACTION:** Any symptomatic pulmonary embolism (PE), fatal PE, symptomatic deep venous thrombosis, all-cause mortality, and major bleeding. Pooled relative risks and associated 95% CIs were calculated. For treatment effects that were statistically significant, the authors determined the absolute risk reduction and the number needed to treat for benefit (NNT(B)) to prevent an outcome. **DATA SYNTHESIS:** 9 studies (n = 19 958) were included. During anticoagulant prophylaxis, patients had significant reductions in any PE (relative risk, 0.43 [CI, 0.26 to 0.71]; absolute risk reduction, 0.29%; NNT(B), 345) and fatal PE (relative risk, 0.38 [CI, 0.21 to 0.69]; absolute risk reduction, 0.25%; NNT(B), 400), a nonsignificant reduction in symptomatic deep venous thrombosis (relative risk, 0.47 [CI, 0.22 to 1.00]), and a nonsignificant increase in major bleeding (relative risk, 1.32 [CI, 0.73 to 2.37]). Anticoagulant prophylaxis had no effect on all-cause mortality (relative risk, 0.97 [CI, 0.79 to 1.19]). **LIMITATIONS:** 2 of 9 included studies were not double-blind. **CONCLUSIONS:** Anticoagulant prophylaxis is effective in preventing symptomatic venous thromboembolism during anticoagulant prophylaxis in at-risk hospitalized medical patients. Additional research is needed to determine the risk for venous thromboembolism in these patients after prophylaxis has been stopped.

Question: Therapy

Study Type: Meta-analysis

Learner Level: Intermediate

Notes: Paper of importance to internists, hospitalists, house officers and anyone who takes care of patients in the hospital. Clear methodology with excellent forest plots for teaching. Can use to illustrate a paper where results do not display any heterogeneity.

Devereaux, P. J., et al. (2005). "How strong is the evidence for the use of perioperative beta blockers in non-cardiac surgery? Systematic review and meta-analysis of randomised controlled trials." *BMJ* 331(7512): 313-321.

OBJECTIVE: To determine the effect of perioperative beta blocker treatment in patients having non-cardiac surgery. **DESIGN:** Systematic review and meta-analysis. **DATA SOURCES:** Seven search strategies, including searching two bibliographic databases and hand searching seven medical journals. **STUDY SELECTION AND OUTCOMES:** We included randomised controlled trials that evaluated beta blocker treatment in patients having non-cardiac surgery. Perioperative outcomes within 30 days of surgery included total mortality, cardiovascular mortality, non-fatal myocardial infarction, non-fatal cardiac arrest, non-fatal stroke, congestive heart failure, hypotension needing treatment, bradycardia needing treatment, and bronchospasm. **RESULTS:** Twenty two trials that randomised a total of 2437 patients met the eligibility criteria. Perioperative beta blockers did not show any statistically significant beneficial effects on any of the individual outcomes and the only nominally statistically significant beneficial relative risk was 0.44 (95% confidence interval 0.20 to 0.97, 99% confidence interval 0.16 to 1.24) for the composite outcome of cardiovascular mortality, non-fatal myocardial infarction, and non-fatal cardiac arrest. Methods adapted from formal interim monitoring boundaries applied to cumulative meta-analysis showed that the evidence failed, by a considerable degree, to meet standards for forgoing additional studies. The individual safety outcomes in patients treated with perioperative beta blockers showed a relative risk for bradycardia needing treatment of 2.27 (95% CI 1.53 to 3.36, 99% CI 1.36 to 3.80) and a nominally statistically significant relative risk for hypotension needing treatment of 1.27 (95% CI 1.04 to 1.56, 99% CI 0.97 to 1.66). **CONCLUSION:** The evidence that perioperative beta blockers reduce major cardiovascular events is encouraging but too unreliable to allow definitive conclusions to be drawn.

Question: Therapy

Study Type: Meta-analysis

Learner Level: Intermediate

Notes: Extremely rigorous methods clearly reported make this a great teaching paper. In addition, results are counter to current practice of broadly applying b-blockers in the perioperative period to patient undergoing non-cardiac surgery. Results are good for teaching about heterogeneity. The section in the results section on exploring heterogeneity can be very instructive.

Ding, M., et al. (2014). "Caffeinated and decaffeinated coffee consumption and risk of type 2 diabetes: a systematic review and a dose-response meta-analysis." *Diabetes Care* 37(2): 569-586.

OBJECTIVE: Previous meta-analyses identified an inverse association of coffee consumption with the risk of type 2 diabetes. However, an updated meta-analysis is needed because new studies comparing the trends of association for caffeinated and decaffeinated coffee have since been published. **RESEARCH DESIGN AND METHODS:** PubMed and Embase were searched for cohort or nested case-control studies that assessed the relationship of coffee consumption and risk of type 2 diabetes from 1966 to February 2013. A restricted cubic spline random-effects model was used. **RESULTS:** Twenty-eight prospective studies were included in the analysis, with 1,109,272 study participants and 45,335 cases of type 2 diabetes. The follow-up duration ranged from 10 months to 20 years. Compared with no or rare coffee consumption, the relative risk (RR; 95% CI) for diabetes was 0.92 (0.90-0.94), 0.85 (0.82-0.88), 0.79 (0.75-0.83), 0.75 (0.71-0.80), 0.71 (0.65-0.76), and 0.67 (0.61-0.74) for 1-6 cups/day, respectively. The RR of diabetes for a 1 cup/day increase was 0.91 (0.89-0.94) for caffeinated coffee consumption and 0.94 (0.91-0.98) for decaffeinated coffee consumption (P for difference = 0.17). **CONCLUSIONS:** Coffee consumption was inversely associated with the risk of type 2 diabetes in a dose-response manner. Both caffeinated and decaffeinated coffee was associated with reduced diabetes risk.

Question: Harm

Study Type: Systematic Review

Learner Level: Intermediate

Notes: Good for teaching Forest Plots and Heterogeneity.

Drummond, M. B., et al. (2008). "Inhaled corticosteroids in patients with stable chronic obstructive pulmonary disease: a systematic review and meta-analysis." *JAMA* 300(20): 2407-2416.

CONTEXT: Recent studies of inhaled corticosteroid (ICS) therapy for managing stable chronic obstructive pulmonary disease (COPD) have yielded conflicting results regarding survival and risk of adverse events. **OBJECTIVE:** To systematically review and quantitatively synthesize the effects of ICS therapy on mortality and adverse events in patients with stable COPD. **DATA SOURCES:** Search of MEDLINE, CENTRAL, EMBASE, CINAHL, Web of Science, and PsychInfo through February 9, 2008. **STUDY SELECTION:** Eligible studies were double-blind, randomized controlled trials comparing ICS therapy for 6 or more months with nonsteroid inhaled therapy in patients with COPD. **DATA EXTRACTION:** Two authors independently abstracted data including study characteristics, all-cause mortality, pneumonia, and bone fractures. The I(2) statistic was used to assess heterogeneity. Study-level data were pooled using a random-effects model (when I(2) > or = 50%) or a fixed-effects model (when I(2) < 50%). For the primary outcome of all-cause mortality at 1 year, our meta-analysis was powered to detect a 1.0% absolute difference in mortality, assuming a 2-sided alpha of .05 and power of 0.80. **RESULTS:** Eleven eligible randomized controlled trials (14,426 participants) were included. In trials with mortality data, no difference was observed in 1-year all-cause mortality (128 deaths among 4636 patients in the treatment group and 148 deaths among 4597 patients in the control group; relative risk [RR], 0.86; 95% confidence interval [CI], 0.68-1.09; P = .20; I(2) = 0%). In the trials with data on pneumonia, ICS therapy was associated with a significantly higher incidence of pneumonia (777 cases among 5405 patients in the treatment group and 561 cases among 5371 patients in the control group; RR, 1.34; 95% CI, 1.03-1.75; P = .03; I(2) = 72%). Subgroup analyses indicated an increased risk of pneumonia in the following subgroups: highest ICS dose (RR, 1.46; 95% CI, 1.10-1.92; P = .008; I(2) = 78%), shorter duration of ICS use (RR, 2.12; 95% CI, 1.47-3.05; P < .001; I(2) = 0%), lowest baseline forced expiratory volume in the first second of expiration (RR, 1.90; 95% CI, 1.26-2.85; P = .002; I(2) = 0%), and combined ICS and bronchodilator therapy (RR, 1.57; 95% CI, 1.35-1.82; P < .001; I(2) = 24%). **CONCLUSIONS:** Among patients with COPD, ICS therapy does not affect 1-year all-cause mortality. ICS therapy is associated with a higher risk of pneumonia. Future studies should determine whether specific subsets of patients with COPD benefit from ICS therapy.

Question: Therapy

Study Type: Meta-analysis

Learner Level: Intermediate to Advanced

Notes: Clearly reported meta-analysis on a somewhat controversial topic (inhaled steroids for COPD) with a controversial result (no change in mortality but increased pneumonia in treated group). This is also a good paper for discussing heterogeneity: mortality outcome in figure 2 (I2=0), but pneumonia risk has significant heterogeneity indicating that it might not be okay to combine (I2=72%). This paper should be reserved for more advanced learners who are ready to grapple with more difficult issues. There is an ACP Journal Club summary of this paper, but remarkably they don't mention the heterogeneity issues.

Eidelman, R. S., et al. (2004). "Randomized trials of vitamin E in the treatment and prevention of cardiovascular disease." *Arch Intern Med* 164(14): 1552-1556.

BACKGROUND: Observational epidemiological studies consistently show that individuals who choose to take high amounts of vitamin E through diet or supplements experience cardiovascular benefits, for which basic research provides plausible mechanisms. However, because the size of the postulated benefit is small to moderate, the confounding inherent in observational studies is as great as the effect size. Before the availability of randomized evidence, about 1 in 4 adults was taking vitamin E supplements in the United States. **METHODS:** We conducted a computerized search of the English-language literature from 1990 to the present and found 7 large-scale randomized trials of the effectiveness vitamin E in the treatment and prevention of cardiovascular disease. Data were available on myocardial infarction, stroke, or cardiovascular death. **RESULTS:** Six of the 7 trials showed no significant effect of

vitamin E on cardiovascular disease. In an overview, vitamin E had neither a statistically significant nor a clinically important effect on any important cardiovascular event (odds ratio [OR], 0.98; 95% confidence interval [CI], 0.94-1.03) or its components: nonfatal myocardial infarction (OR, 1.00; 95% CI, 0.92-1.09), nonfatal stroke (OR, 1.03; 95% CI, 0.93-1.14), or cardiovascular death (OR, 1.00; 95% CI, 0.94-1.05). **CONCLUSIONS:** The ORs and CIs provide strong support for a lack of statistically significant or clinically important effects of vitamin E on cardiovascular disease. The use of agents of proven lack of benefit, especially those easily available over the counter, may contribute to underuse of agents of proven benefit and failure to adopt healthy lifestyles.

Question: Therapy

Study Type: Systematic Review

Learner Level: Intermediate

Notes: Very poor quality systematic review. Give very little information on methods. Good Discussion Points: This is an excellent article to use in combination with the Shekelle article to highlight the differences in quality. This is a good example of inappropriate pooling of studies. Pooling combined antioxidants with vitamin E alone. Good example of how poor quality articles can be published in well regarded journals. Can be used to force learners to think about how they would have done the review.

Feller, M., et al. (2018). "Association of Thyroid Hormone Therapy With Quality of Life and Thyroid-Related Symptoms in Patients With Subclinical Hypothyroidism: A Systematic Review and Meta-analysis." *JAMA* 320(13): 1349-1359.

Importance: The benefit of thyroid hormone therapy for subclinical hypothyroidism is uncertain. New evidence from recent large randomized clinical trials warrants an update of previous meta-analyses. **Objective:** To conduct a meta-analysis of the association of thyroid hormone therapy with quality of life and thyroid-related symptoms in adults with subclinical hypothyroidism. **Data Sources:** PubMed, EMBASE, ClinicalTrials.gov, Web of Science, Cochrane Library, CENTRAL, Emcare, and Academic Search Premier from inception until July 4, 2018. **Study Selection:** Randomized clinical trials that compared thyroid hormone therapy with placebo or no therapy in nonpregnant adults with subclinical hypothyroidism were eligible. Two reviewers independently evaluated eligibility based on titles and abstracts of all retrieved studies. Studies not excluded in this first step were independently assessed for inclusion after full-text evaluation by 2 reviewers. **Data Extraction and Synthesis:** Two independent reviewers extracted data, assessed risk of bias (Cochrane risk-of-bias tool), and evaluated the quality of evidence (GRADE tool). For synthesis, differences in clinical scores were transformed (eg, quality of life) into standardized mean differences (SMDs; positive values indicate benefit of thyroid hormone therapy; 0.2, 0.5, and 0.8 correspond to small, moderate, and large effects, respectively). Random-effects models for meta-analyses were applied. **Main Outcomes and Measures:** General quality of life and thyroid-related symptoms after a minimum follow-up of 3 months. **Results:** Overall, 21 of 3088 initially identified publications met the inclusion criteria, with 2192 adults randomized. After treatment (range, 3-18 months), thyroid hormone therapy was associated with lowering the mean thyrotropin value into the normal reference range compared with placebo (range, 0.5-3.7 mIU/L vs 4.6 to 14.7 mIU/L) but was not associated with benefit regarding general quality of life (n = 796; SMD, -0.11; 95% CI, -0.25 to 0.03; I²=66.7%) or thyroid-related symptoms (n = 858; SMD, 0.01; 95% CI, -0.12 to 0.14; I²=0.0%). Overall, risk of bias was low and the quality of evidence assessed with the GRADE tool was judged moderate to high. **Conclusions and Relevance:** Among nonpregnant adults with subclinical hypothyroidism, the use of thyroid hormone therapy was not associated with improvements in general quality of life or thyroid-related symptoms. These findings do not support the routine use of thyroid hormone therapy in adults with subclinical hypothyroidism.

Question: Therapy

Study Type: Systematic Review - Meta-Analysis

Learner Level: Intermediate

Notes: A nicely reported SR and Meta-analysis on a common topic in internal medicine. There are good Forest plots for teaching and a good take home message that will lend itself to an interactive teaching scenario (to treat or not to treat?). There is an ACP journal club for interested learners and new teachers of systematic reviews / meta-analysis.

Hlibczuk, V., et al. (2010). "Diagnostic accuracy of noncontrast computed tomography for appendicitis in adults: a systematic review." *Ann Emerg Med* 55(1): 51-59 e51.

STUDY OBJECTIVE: We seek to determine the diagnostic test characteristics of noncontrast computed tomography (CT) for appendicitis in the adult emergency department (ED) population. **METHODS:** We conducted a search of MEDLINE, EMBASE, the Cochrane Library, and the bibliographies of previous systematic reviews. Included studies assessed the diagnostic accuracy of noncontrast CT for acute appendicitis in adults by using the final diagnosis at surgery or follow-up at a minimum of 2 weeks as the reference standard. Studies were included only if the CT was completed using a multislice helical scanner. Two authors independently conducted the relevance screen of titles and abstracts, selected studies for the final inclusion, extracted data, and assessed study quality. Consensus was reached by conference, and any disagreements were adjudicated by a third reviewer. Unenhanced CT test performance was assessed with summary receiver operating characteristic curve analysis, with independently pooled sensitivity and specificity values across studies. **RESULTS:** The search yielded 1,258 publications; 7 studies met the inclusion criteria and provided a sample of 1,060 patients. The included studies were of high methodological quality with respect to appropriate patient spectrum and reference standard. Our pooled estimates for sensitivity and specificity were 92.7% (95% confidence interval 89.5% to 95.0%) and 96.1% (95% confidence interval 94.2% to 97.5%), respectively; the positive likelihood ratio=24 and the negative likelihood ratio=0.08. **CONCLUSION:** We found the diagnostic accuracy of noncontrast CT for the diagnosis of acute appendicitis in the adult population to be adequate for clinical decisionmaking in the ED setting.

Question: Diagnosis

Study Type: Systematic Review

Learner Level: .

Notes: Non-contrast appi. While most people think of SRs as involving therapy issues, there are also diagnosis SRs which offer some unique challenges from a critical appraisal perspective. This article can also serve as great substrate for a GRADE exercise or a Knowledge Translation effort.

Hollingsworth, J. M., et al. (2006). "Medical therapy to facilitate urinary stone passage: a meta-analysis." *Lancet* 368(9542): 1171-1179.

BACKGROUND: Medical therapies to ease urinary-stone passage have been reported, but are not generally used. If effective, such therapies would increase the options for treatment of urinary stones. To assess efficacy, we sought to identify and summarise all randomised controlled trials in which calcium-channel blockers or alpha blockers were used to treat urinary stone disease. **METHODS:** We searched MEDLINE, Pre-MEDLINE, CINAHL, and EMBASE, as well as scientific meeting abstracts, up to July, 2005. All randomised controlled trials in which calcium-channel blockers or alpha blockers were used to treat ureteral stones were eligible for inclusion in our analysis. Data from nine trials (number of patients=693) were pooled. The main outcome was the proportion of patients who passed stones. We calculated the summary estimate of effect associated with medical therapy use using random-effects and fixed-effects models. **FINDINGS:** Patients given calcium-channel blockers or alpha blockers had a 65% (absolute risk reduction=0.31 95% CI 0.25-0.38) greater likelihood of stone passage than those not given such treatment (pooled risk ratio 1.65; 95% CI 1.45-1.88). The pooled risk ratio for alpha blockers was 1.54 (1.29-1.85) and for calcium-channel blockers with steroids was 1.90 (1.51-2.40). The proportion of heterogeneity not explained by chance alone was 28%. The number needed to treat was 4. **INTERPRETATION:** Although a high-quality randomised trial is necessary to confirm its efficacy, our findings suggest that medical therapy is an option for facilitation of urinary-stone passage for patients amenable to conservative management, potentially obviating the need for surgery.

Question: Therapy

Study Type: Meta-analysis

Learner Level: Beginner

Notes: The study pools relatively few studies of low methodological quality. Suitable for group exercises in which learners abstract data assess methodological quality of individual RCTS and then pool data.

Horita, N., et al. (2017). "Long-acting muscarinic antagonist (LAMA) plus long-acting beta-agonist (LABA) versus LABA plus inhaled corticosteroid (ICS) for stable chronic obstructive pulmonary disease (COPD)." *Cochrane Database Syst Rev* 2: Cd012066.

BACKGROUND: Three classes of inhaler medications are used to manage chronic obstructive pulmonary disease (COPD): long-acting beta-agonists (LABA), long-acting muscarinic antagonists (LAMA), and inhaled corticosteroids (ICS). When two classes of medications are required, LAMA plus LABA (LAMA+LABA) and LABA plus ICS (LABA+ICS) are often selected because these combinations can be administered via a single medication device. The previous Global Initiative for Chronic Obstructive Lung Disease (GOLD) guidance recommended LABA+ICS as the first-line treatment for managing stable COPD in high-risk people of categories C and D. However, the updated GOLD 2017 guidance recommends LAMA+LABA over LABA+ICS. **OBJECTIVES:** To compare the benefits and harms of LAMA+LABA versus LABA+ICS for treatment of people with stable COPD. **SEARCH METHODS:** We performed an electronic search of the Cochrane Airways Group Specialised Register (2 February 2016), ClinicalTrials.gov (4 June 2016), and the World Health Organization Clinical Trials Search Portal (4 June 2016), followed by a handsearch (5 June 2016). Two review authors screened and scrutinised the selected articles. **SELECTION CRITERIA:** We included individual randomised controlled trials, parallel-group trials, and cross-over trials comparing LAMA+LABA and LABA+ICS for stable COPD. The minimum accepted trial duration was one month and trials should have been conducted in an outpatient setting. **DATA COLLECTION AND ANALYSIS:** Two review authors independently extracted data and evaluated risk of bias. We resolved any discrepancies through discussion. We analysed dichotomous data as odds ratios (OR), and continuous data as mean differences (MD), with 95% confidence interval (CI) using Review Manager 5. Exacerbations were measured by counting the number of people experiencing one or more exacerbation. **MAIN RESULTS:** We included 11 studies comprising 9839 participants in our quantitative analysis. Most studies included people with moderate to severe COPD, without recent exacerbations. One pharmaceutical sponsored trial that included only people with recent exacerbations was the largest study and accounted for 37% of participants. All but one study were sponsored by pharmaceutical companies, thus we rated them as having a high risk of 'other bias'. The unsponsored study was at high risk of performance and detection bias, and possible selective reporting. Five studies recruited GOLD Category B participants, one study recruited Category D participants, two studies recruited Category A/B participants, and three studies recruited participants regardless of category. Follow-up ranged from 6 to 52 weeks. Compared to the LABA+ICS arm, the results for the pooled primary outcomes for the LAMA+LABA arm were as follows: exacerbations, OR 0.82 (95% CI 0.70 to 0.96, P = 0.01, I(2) = 17%, low quality evidence); serious adverse events (SAE), OR 0.91 (95% CI 0.79 to 1.05, P = 0.18, I(2) = 0, moderate quality evidence); St. George's Respiratory Questionnaire (SGRQ) total score change from the baseline, MD -1.22 (95% CI -2.52 to 0.07, P = 0.06, I(2) = 71%, low quality evidence); and trough forced expiratory volume in one second (FEV1) change from the baseline, MD 0.08 L (95% CI 0.06 to 0.09, P < 0.0001, I(2) = 50%, moderate quality evidence). Compared to the LABA+ICS arm, the results for the pooled secondary outcomes for the LAMA+LABA arm were as follows: pneumonia, OR 0.57 (95% CI 0.42 to 0.79, P = 0.0006, I(2) = 0%, low quality evidence); all-cause death, OR 1.01 (95% CI 0.61 to 1.67, P = 0.88, I(2) = 0%, low quality evidence); and SGRQ total score change from the baseline of 4 points or greater (the minimal clinically important difference for the SGRQ is 4 points), OR 1.25 (95% CI 1.09 to 1.44, P = 0.002, I(2) = 0%, moderate quality evidence). **AUTHORS' CONCLUSIONS:** For the treatment of COPD, LAMA+LABA has fewer exacerbations, a larger improvement of FEV1, a lower

risk of pneumonia, and more frequent improvement in quality of life as measured by an increase over 4 units or more of the SGRO. These data were supported by low or moderate quality evidence generated from mainly participants with moderate to severe COPD in heterogeneous trials with an observation period of less than one year. Our findings support the recently updated GOLD guidance.

Question: Therapy

Study Type: Systematic Review

Learner Level: Intermediate

Notes: Good for teaching Forest Plots and Heterogeneity.

Hull, R. D., et al. (2001). "Extended out-of-hospital low-molecular-weight heparin prophylaxis against deep venous thrombosis in patients after elective hip arthroplasty: a systematic review." Ann Intern Med 135(10): 858-869.

PURPOSE: Evidence-based medicine guidelines based on venographic end points recommend in-hospital prophylaxis with low-molecular-weight heparin (LMWH) in patients having elective hip surgery. Emerging data suggest that out-of-hospital use may offer additional protection; however, uncertainty remains about the risk-benefit ratio. To provide clinicians with a practical pathway for translating clinical research into practice, we systematically reviewed trials comparing extended out-of-hospital LMWH prophylaxis versus placebo. **DATA SOURCES:** Studies were identified by 1) searching PubMed, MEDLINE, and the Cochrane Library Database for reports published from January 1976 to May 2001; 2) reviewing references from retrieved articles; 3) scanning abstracts from conference proceedings; and 4) contacting pharmaceutical companies and investigators of the original reports. **STUDY SELECTION:** Randomized, controlled trials comparing extended out-of-hospital prophylaxis with LMWH versus placebo in patients having elective hip arthroplasty. **DATA EXTRACTION:** Two reviewers extracted data independently. Reviewers evaluated study quality by using a validated four-item instrument. **DATA SYNTHESIS:** Six of seven original articles met the defined inclusion criteria. The included studies were double-blind trials that used proper randomization procedures. Compared with placebo, extended out-of-hospital prophylaxis decreased the frequency of all episodes of deep venous thrombosis (placebo rate, 150 of 666 patients [22.5%]; relative risk, 0.41 [95% CI, 0.32 to 0.54; $P < 0.001$]), proximal venous thrombosis (placebo rate, 76 of 678 patients [11.2%]; relative risk, 0.31 [CI, 0.20 to 0.47; $P < 0.001$]), and symptomatic venous thromboembolism (placebo rate, 36 of 862 patients [4.2%]; relative risk, 0.36 [CI, 0.20 to 0.67; $P = 0.001$]). Major bleeding was rare, occurring in only one patient in the placebo group. **CONCLUSIONS:** Extended LMWH prophylaxis showed consistent effectiveness and safety in the trials (regardless of study variations in clinical practice and length of hospital stay) for venographic deep venous thrombosis and symptomatic venous thromboembolism. The aggregate findings support the need for extended out-of-hospital prophylaxis in patients undergoing hip arthroplasty surgery.

Question: Therapy

Study Type: Meta-analysis

Learner Level: Intermediate

Notes: Important question which prompts good discussion of risks v. benefits v. cost. Good Discussion Points: Strengths: Well designed with clear description of methods. Can use methods to prompt vocabulary discussion for concepts pertaining to Meta-analysis (e.g. methods for validity assessment, summary treatment effects, sensitivity analysis, heterogeneity); Weaknesses: Variety of interventions can become very confusing to learners. More Advanced Discussion: might consider getting the QUOROM statement that is referenced in the methods of the paper (p. 859) for more in depth discussion of Meta-analysis methodology. Check list on p 1897 can be used to assess this (or any other systematic review); For more advanced discussion see also Moher 1999.

Humphrey, L. L., et al. (2002). "Postmenopausal hormone replacement therapy and the primary prevention of cardiovascular disease." Ann Intern Med 137(4): 273-284.

PURPOSE: To evaluate the value of hormone replacement therapy (HRT) in the primary prevention of cardiovascular disease (CVD) and coronary artery disease (CAD). **DATA SOURCES:** MEDLINE and Cochrane databases were searched for all primary prevention studies reporting CVD or CAD incidence, mortality, or both in association with HRT; reference lists, letters, editorials, and reviews were also reviewed. **DATA EXTRACTION:** All studies were reviewed, abstracted, and rated for quality. **STUDY SELECTION:** Only studies of good or fair quality, according to U.S. Preventive Services Task Force (USPSTF) criteria, were included in the detailed review and meta-analysis. **DATA SYNTHESIS:** The summary relative risk with any HRT use was 0.75 (95% credible interval [CrI], 0.42 to 1.23) for CVD mortality and 0.74 (CrI, 0.36 to 1.45) for CAD mortality. The summary relative risk with any use was 1.28 (CrI, 0.86 to 2.00) for CVD incidence and 0.87 (CrI, 0.62 to 1.21) for CAD incidence. Further analysis of studies adjusting for socioeconomic status, as well as other major CAD risk factors, showed a summary relative risk of 1.07 (CrI, 0.79 to 1.48) for CAD incidence associated with any HRT use. Similar results were found when the analysis was stratified by studies adjusting for alcohol consumption, exercise, or both, in addition to other major risk factors, suggesting confounding by these factors. **CONCLUSIONS:** This meta-analysis differs from previous meta-analyses by evaluating potential explanatory variables of the relationship between HRT, CVD, and CAD. The adjusted meta-analysis is consistent with recent randomized trials that have shown no benefit in the secondary or primary prevention of CVD events. A valid answer to the role of HRT in the primary prevention of CVD will best come from randomized, controlled trials.

Question: Therapy

Study Type: Meta-analysis

Learner Level: Intermediate

Notes: Pros: Can discuss Cochrane controlled trials register. Validity criteria for inclusion of studies into the SR. Search strategies for SRs. Cons: Long, Dated

Huss, A., et al. (2009). "Efficacy of pneumococcal vaccination in adults: a meta-analysis." *CMAJ* 180(1): 48-58.

BACKGROUND: Clinical trials and meta-analyses have produced conflicting results of the efficacy of unconjugated pneumococcal polysaccharide vaccine in adults. We sought to evaluate the vaccine's efficacy on clinical outcomes as well as the methodologic quality of the trials. **METHODS:** We searched several databases and all bibliographies of reviews and meta-analyses for clinical trials that compared pneumococcal polysaccharide vaccine with a control. We examined rates of pneumonia and death, taking the methodologic quality of the trials into consideration. **RESULTS:** We included 22 trials involving 101 507 participants: 11 trials reported on presumptive pneumococcal pneumonia, 19 on all-cause pneumonia and 12 on all-cause mortality. The current 23-valent vaccine was used in 8 trials. The relative risk (RR) was 0.64 (95% confidence interval [CI] 0.43-0.96) for presumptive pneumococcal pneumonia and 0.73 (95% CI 0.56-0.94) for all-cause pneumonia. There was significant heterogeneity between the trials reporting on presumptive pneumonia ($I(2) = 74\%$, $p < 0.001$) and between those reporting on all-cause pneumonia ($I(2) = 90\%$, $p < 0.001$). The RR for all-cause mortality was 0.97 (95% CI 0.87-1.09), with moderate heterogeneity between trials ($I(2) = 44\%$, $p = 0.053$). Trial quality, especially regarding double blinding, explained a substantial proportion of the heterogeneity in the trials reporting on presumptive pneumonia and all-cause pneumonia. There was little evidence of vaccine protection in trials of higher methodologic quality (RR 1.20, 95% CI 0.75-1.92, for presumptive pneumonia; and 1.19, 95% CI 0.95-1.49, for all-cause pneumonia in double-blind trials; p for heterogeneity > 0.05). The results for all-cause mortality in double-blind trials were similar to those in all trials combined. There was little evidence of vaccine protection among elderly patients or adults with chronic illness in analyses of all trials (RR 1.04, 95% CI 0.78-1.38, for presumptive pneumococcal pneumonia; 0.89, 95% CI 0.69-1.14, for all-cause pneumonia; and 1.00, 95% CI 0.87-1.14, for all-cause mortality). **INTERPRETATION:** Pneumococcal vaccination does not appear to be effective in preventing pneumonia, even in populations for whom the vaccine is currently recommended.

Question: Therapy

Study Type: Meta-analysis

Learner Level: Beginner

Notes: article is suited to describe the methodology, advantages and limitations of a meta-analysis. Also useful for the purposes of recognizing the impact of primary trial heterogeneity on the results of a meta-analysis.

James, P. A., et al. (2014). "2014 evidence-based guideline for the management of high blood pressure in adults: report from the panel members appointed to the Eighth Joint National Committee (JNC 8)." *JAMA* 311(5): 507-520.

Hypertension is the most common condition seen in primary care and leads to myocardial infarction, stroke, renal failure, and death if not detected early and treated appropriately. Patients want to be assured that blood pressure (BP) treatment will reduce their disease burden, while clinicians want guidance on hypertension management using the best scientific evidence. This report takes a rigorous, evidence-based approach to recommend treatment thresholds, goals, and medications in the management of hypertension in adults. Evidence was drawn from randomized controlled trials, which represent the gold standard for determining efficacy and effectiveness. Evidence quality and recommendations were graded based on their effect on important outcomes. There is strong evidence to support treating hypertensive persons aged 60 years or older to a BP goal of less than 150/90 mm Hg and hypertensive persons 30 through 59 years of age to a diastolic goal of less than 90 mm Hg; however, there is insufficient evidence in hypertensive persons younger than 60 years for a systolic goal, or in those younger than 30 years for a diastolic goal, so the panel recommends a BP of less than 140/90 mm Hg for those groups based on expert opinion. The same thresholds and goals are recommended for hypertensive adults with diabetes or nondiabetic chronic kidney disease (CKD) as for the general hypertensive population younger than 60 years. There is moderate evidence to support initiating drug treatment with an angiotensin-converting enzyme inhibitor, angiotensin receptor blocker, calcium channel blocker, or thiazide-type diuretic in the nonblack hypertensive population, including those with diabetes. In the black hypertensive population, including those with diabetes, a calcium channel blocker or thiazide-type diuretic is recommended as initial therapy. There is moderate evidence to support initial or add-on antihypertensive therapy with an angiotensin-converting enzyme inhibitor or angiotensin receptor blocker in persons with CKD to improve kidney outcomes. Although this guideline provides evidence-based recommendations for the management of high BP and should meet the clinical needs of most patients, these recommendations are not a substitute for clinical judgment, and decisions about care must carefully consider and incorporate the clinical characteristics and circumstances of each individual patient.

Question: Therapy

Study Type: Guideline

Learner Level: Beginner

Notes: Type: Clinical Practice Guideline. Teaching points: important, well done, evidence based guideline, transparently reported. The results are controversial as there was not agreement on the panel for recommendations for BP cut offs in individuals 60 or older (without DM or renal disease). The nice part about this study is that it makes very clear links to RCT-level evidence and limits its discussion and recommendations to that which can be clearly linked to high quality evidence. Learners can discuss the pluses and minuses of that approach. As an extra added benefit: it is very hard to find a CPG that is concise enough to print for teaching and this one is.

Juni, P., et al. (2004). "Risk of cardiovascular events and rofecoxib: cumulative meta-analysis." *Lancet* 364(9450): 2021-2029.

BACKGROUND: The cyclo-oxygenase 2 inhibitor rofecoxib was recently withdrawn because of cardiovascular adverse effects. An increased risk of myocardial infarction had been observed in 2000 in the Vioxx Gastrointestinal Outcomes Research study (VIGOR), but was attributed to cardioprotection of naproxen rather than a cardiotoxic effect of rofecoxib. We used standard and cumulative random-effects meta-analyses of randomised controlled trials and observational studies to establish whether robust evidence on the adverse effects of rofecoxib was available before September, 2004. **METHODS:** We searched bibliographic databases and relevant files of the US Food and Drug Administration. We included all randomised controlled trials in patients with chronic musculoskeletal disorders that compared rofecoxib with other non-steroidal anti-inflammatory drugs (NSAIDs) or placebo, and cohort and case-control studies of cardiovascular risk and naproxen. Myocardial infarction was the primary endpoint. **FINDINGS:** We identified 18 randomised controlled trials and 11 observational studies. By the end of 2000 (52 myocardial infarctions, 20742 patients) the relative risk from randomised controlled trials was 2.30 (95% CI 1.22-4.33, $p=0.010$), and 1 year later (64 events, 21432 patients) it was 2.24 (1.24-4.02, $p=0.007$). There was little evidence that the relative risk differed depending on the control group (placebo, non-naproxen NSAID, or naproxen; $p=0.41$) or trial duration ($p=0.82$). In observational studies, the cardioprotective effect of naproxen was small (combined estimate 0.86 [95% CI 0.75-0.99]) and could not have explained the findings of the VIGOR trial. **INTERPRETATION:** Our findings indicate that rofecoxib should have been withdrawn several years earlier. The reasons why manufacturer and drug licensing authorities did not continuously monitor and summarise the accumulating evidence need to be clarified.

Question: Therapy (Harm)

Study Type: Meta-analysis

Learner Level: Intermediate

Notes: Topical question of broad clinical interest which learners find interesting. Robust cumulative meta-analysis showing the chronology of relevant RCTs on the topic demonstrating harm of this commonly prescribed medication.

Keeney, G. E., et al. (2014). "Dexamethasone for acute asthma exacerbations in children: a meta-analysis." *Pediatrics* 133(3): 493-499.

BACKGROUND AND OBJECTIVE: Dexamethasone has been proposed as an equivalent therapy to prednisone/prednisolone for acute asthma exacerbations in pediatric patients. Although multiple small trials exist, clear consensus data are lacking. This systematic review and meta-analysis aimed to determine whether intramuscular or oral dexamethasone is equivalent or superior to a 5-day course of oral prednisone or prednisolone. The primary outcome of interest was return visits or hospital readmissions. **METHODS:** A search of PubMed (Medline) through October 19, 2013, by using the keywords dexamethasone or decadron and asthma or status asthmaticus identified potential studies. Six randomized controlled trials in the emergency department of children ≤ 18 years of age comparing dexamethasone with prednisone/prednisolone for the treatment of acute asthma exacerbations were included. Data were abstracted by 4 authors and verified by a second author. Two reviewers evaluated study quality independently and interrater agreement was assessed. **RESULTS:** There was no difference in relative risk (RR) of relapse between the 2 groups at any time point (5 days RR 0.90, 95% confidence interval [CI] 0.46-1.78, $Q = 1.86$, $df = 3$, $I^2 = 0.0\%$, 10-14 days RR 1.14, 95% CI 0.77-1.67, $Q = 0.84$, $df = 2$, $I^2 = 0.0\%$, or 30 days RR 1.20, 95% CI 0.03-56.93). Patients who received dexamethasone were less likely to experience vomiting in either the emergency department (RR 0.29, 95% CI 0.12-0.69, $Q = 3.78$, $df = 3$, $I^2 = 20.7\%$) or at home (RR 0.32, 95% CI 0.14-0.74, $Q = 2.09$, $df = 2$, $I^2 = 4.2\%$). **CONCLUSIONS:** Practitioners should consider single or 2-dose regimens of dexamethasone as a viable alternative to a 5-day course of prednisone/prednisolone.

Question: Therapy

Study Type: Meta-Analysis

Learner Level: Beginner/Intermediate

Notes: Not the greatest meta-analysis overall (with regards to should these article be combined) but has good F's: frame it, find it, filter/fetch it, forrest plots. Can have good discussion regarding the ultimate analysis of the results – should these studies have been combined and analyzed by meta-analysis or simply reported as a systematic review? 6 studies – 3 used intramuscular dexamethasone and 3 used oral. Of the 3 with oral dex, 2 used 1 daily dose on 2 consecutive days (1 in ED and 1 at home) while the other used a single dose in the ED. The doses and duration of the "control" which was prednisolone were different in each study.

Liu, Z., et al. (2009). "Carotid artery stenting versus carotid endarterectomy: systematic review and meta-analysis." *World J Surg* 33(3): 586-596.

BACKGROUND: The aim of this study was to evaluate and investigate the complications of carotid endarterectomy (CEA) and carotid artery stenting (CAS) by performing a meta-analysis based on prospective randomized controlled trials (RCTs). **METHODS:** We performed a search of multiple electronic databases for RCTs containing patients with carotid stenosis who underwent CAS or CEA, focusing on studies published during 1995-2008. **RESULTS:** Eight trials with 2942 patients (1462 with CEA, 1480 with CAS) were analyzed. The pooled relative risk (RR) after CEA for stroke/death 30 days or 1 year was similar to that for CAS. Thirty-day RR = 0.69, 95% confidence interval (CI) = 0.45-1.07, $p = 0.10$. One-year RR = 0.88, 95% CI = 0.43-1.79, $p = 0.72$. The rates of death, disabling stroke, and nondisabling stroke at 30 days did not differ significantly between CEA and CAS in the subgroup analysis. Compared with CEA, the relative risk of disabling stroke/death within 30 days was not significantly less for CAS with embolic protection devices (EPDs). The relative risk of myocardial infarction within 30 days, myocardial infarction within 1 year, and cervical/peripheral nerve injury within 30 days were significantly higher after CEA; the relative risk of bradycardia/hypotension within 30 days and the 1-year

restenosis rate were significantly higher after CAS. CONCLUSIONS: CAS is equal to CEA with regard to the incidence of stroke/death. These procedures may be considered complementary rather than competing modes of therapy, each of which can be optimized with careful patient selection. CAS with an EPD may be appropriate in certain patients, and in general CAS should be considered cautiously in symptomatic patients.

Question: Therapy

Study Type: Meta-analysis

Learner Level: Intermediate

Notes: Good to teach heterogeneity

Luther, M. K., et al. (2018). "Vancomycin Plus Piperacillin-Tazobactam and Acute Kidney Injury in Adults: A Systematic Review and Meta-Analysis." *Critical Care Medicine* 46(1): 12-20.

OBJECTIVES: The objective of this systematic review and meta-analysis was to assess acute kidney injury with combination therapy of vancomycin plus piperacillin-tazobactam, in general, adult patients and in critically ill adults. Rates of acute kidney injury, time to acute kidney injury, and odds of acute kidney injury were compared with vancomycin monotherapy, vancomycin plus cefepime or carbapenem, or piperacillin-tazobactam monotherapy. **DATA SOURCES:** Studies were identified by searching Pubmed, Embase, Web of Science, and Cochrane from inception to April 2017. Abstracts from selected conference proceedings were manually searched. **STUDY SELECTION:** Articles not in English, pediatric studies, and case reports were excluded. **DATA EXTRACTION:** Two authors independently extracted data on study methods, rates of acute kidney injury, and time to acute kidney injury. Effect estimates and 95% CIs were calculated using the random effects model in RevMan 5.3. **DATA SYNTHESIS:** Literature search identified 15 published studies and 17 conference abstracts with at least 24,799 patients. The overall occurrence rate of acute kidney injury was 16.7%, with 22.2% for vancomycin plus piperacillin-tazobactam and 12.9% for comparators. This yielded an overall number needed to harm of 11. Time to acute kidney injury was faster for vancomycin plus piperacillin-tazobactam than vancomycin plus cefepime or carbapenem, but not significantly (mean difference, -1.30; 95% CI, -3.00 to 0.41 d). The odds of acute kidney injury with vancomycin plus piperacillin-tazobactam were increased versus vancomycin monotherapy (odds ratio, 3.40; 95% CI, 2.57-4.50), versus vancomycin plus cefepime or carbapenem (odds ratio, 2.68; 95% CI, 1.83-3.91), and versus piperacillin-tazobactam monotherapy (odds ratio, 2.70; 95% CI, 1.97-3.69). In a small subanalysis of 968 critically ill patients, the odds of acute kidney injury were increased versus vancomycin monotherapy (odds ratio, 9.62; 95% CI, 4.48-20.68), but not significantly different for vancomycin plus cefepime or carbapenem (odds ratio, 1.43; 95% CI, 0.83-2.47) or piperacillin-tazobactam monotherapy (odds ratio, 1.35; 95% CI, 0.86-2.11). **CONCLUSIONS:** The combination of vancomycin plus piperacillin-tazobactam increased the odds of acute kidney injury over vancomycin monotherapy, vancomycin plus cefepime or carbapenem, and piperacillin-tazobactam monotherapy. Limited data in critically ill patients suggest the odds of acute kidney injury are increased versus vancomycin monotherapy, and mitigated versus the other comparators. Further research in the critically ill population is needed.

Question: Harm

Study Type: Meta-Analysis

Learner Level: .

Notes: .

McGee, S. R. and E. J. Boyko (1998). "Physical examination and chronic lower-extremity ischemia: a critical review." *Arch Intern Med* 158(12): 1357-1364.

OBJECTIVE: To determine the clinical utility of physical examination in patients with suspected chronic ischemia of the lower extremities. **DATA SOURCES:** MEDLINE search (January 1966 to January 1997), personal files, and bibliographies of textbooks on physical diagnosis, surgery, and vascular surgery. **STUDY SELECTION:** Both authors independently graded the studies as level 1, 2, or 3, according to predetermined criteria. Criteria deemed essential for analysis of sensitivity, specificity, and likelihood ratios were (1) clear definition of study population, (2) clear definition of physical examination maneuver, and (3) use of an acceptable criterion standard test for comparison. **RESULTS:** The following positive findings help clinicians diagnose the presence of peripheral arterial disease: abnormal pedal pulses, a unilaterally cool extremity, a prolonged venous filling time, and a femoral bruit. Other physical signs help determine the extent and distribution of vascular disease, including an abnormal femoral pulse, lower-extremity bruits, warm knees, and the Buerger test. The capillary refill test and the findings of foot discoloration, atrophic skin, and hairless extremities are unhelpful in diagnostic decisions. Mathematical formulas, derived from 2 studies using multivariate analysis, allow clinicians to estimate the probability of peripheral arterial disease in their patients. **CONCLUSION:** Certain aspects of the physical examination help clinicians make accurate judgments about the presence of peripheral arterial disease and its distribution.

Question: Diagnosis

Study Type: Systematic Review

Learner Level: Beginner / Intermediate

Notes: Solid methods with comparison to reference standards. Although this is not a meta-analysis (i.e. they didn't combine results) it is a good systematic review; Good discussion points; Good paper for discussion of diagnosis, kappa (interobserver agreement) and

likelihood ratios as well as systematic review. Down side: because it is not a meta-analysis, you can't discuss certain issues such as heterogeneity

Miller, E. R., 3rd, et al. (2005). "Meta-analysis: high-dosage vitamin E supplementation may increase all-cause mortality." *Ann Intern Med* 142(1): 37-46.

BACKGROUND: Experimental models and observational studies suggest that vitamin E supplementation may prevent cardiovascular disease and cancer. However, several trials of high-dosage vitamin E supplementation showed non-statistically significant increases in total mortality. **PURPOSE:** To perform a meta-analysis of the dose-response relationship between vitamin E supplementation and total mortality by using data from randomized, controlled trials. **PATIENTS:** 135,967 participants in 19 clinical trials. Of these trials, 9 tested vitamin E alone and 10 tested vitamin E combined with other vitamins or minerals. The dosages of vitamin E ranged from 16.5 to 2000 IU/d (median, 400 IU/d). **DATA SOURCES:** PubMed search from 1966 through August 2004, complemented by a search of the Cochrane Clinical Trials Database and review of citations of published reviews and meta-analyses. No language restrictions were applied. **DATA EXTRACTION:** 3 investigators independently abstracted study reports. The investigators of the original publications were contacted if required information was not available. **DATA SYNTHESIS:** 9 of 11 trials testing high-dosage vitamin E (≥ 400 IU/d) showed increased risk (risk difference > 0) for all-cause mortality in comparisons of vitamin E versus control. The pooled all-cause mortality risk difference in high-dosage vitamin E trials was 39 per 10,000 persons (95% CI, 3 to 74 per 10,000 persons; $P = 0.035$). For low-dosage vitamin E trials, the risk difference was -16 per 10,000 persons (CI, -41 to 10 per 10,000 persons; $P > 0.2$). A dose-response analysis showed a statistically significant relationship between vitamin E dosage and all-cause mortality, with increased risk of dosages greater than 150 IU/d. **LIMITATIONS:** High-dosage (≥ 400 IU/d) trials were often small and were performed in patients with chronic diseases. The generalizability of the findings to healthy adults is uncertain. Precise estimation of the threshold at which risk increases is difficult. **CONCLUSION:** High-dosage (≥ 400 IU/d) vitamin E supplements may increase all-cause mortality and should be avoided.

Question: Therapy

Study Type: Meta-analysis

Learner Level: Intermediate

Notes: Solid methodology; Recent, pertinent to many disciplines; Good Discussion Points: Nice illustration of how meta-analysis is used to detect important, but rare adverse events occurring in clinical trials. Results are clearly laid out including data table, forest plots, dose-effect curve; Good paper to discuss confidence intervals (Figures 2 and 4 on pages 42 and 44). The length of the paper may intimidate some earlier learners

Moher, D., et al. (1999). "Improving the quality of reports of meta-analyses of randomised controlled trials: the QUOROM statement. Quality of Reporting of Meta-analyses." *Lancet* 354(9193): 1896-1900.

BACKGROUND: The Quality of Reporting of Meta-analyses (QUOROM) conference was convened to address standards for improving the quality of reporting of meta-analyses of clinical randomised controlled trials (RCTs). **METHODS:** The QUOROM group consisted of 30 clinical epidemiologists, clinicians, statisticians, editors, and researchers. In conference, the group was asked to identify items they thought should be included in a checklist of standards. Whenever possible, checklist items were guided by research evidence suggesting that failure to adhere to the item proposed could lead to biased results. A modified Delphi technique was used in assessing candidate items. **FINDINGS:** The conference resulted in the QUOROM statement, a checklist, and a flow diagram. The checklist describes our preferred way to present the abstract, introduction, methods, results, and discussion sections of a report of a meta-analysis. It is organised into 21 headings and subheadings regarding searches, selection, validity assessment, data abstraction, study characteristics, and quantitative data synthesis, and in the results with "trial flow", study characteristics, and quantitative data synthesis; research documentation was identified for eight of the 18 items. The flow diagram provides information about both the numbers of RCTs identified, included, and excluded and the reasons for exclusion of trials. **INTERPRETATION:** We hope this report will generate further thought about ways to improve the quality of reports of meta-analyses of RCTs and that interested readers, reviewers, researchers, and editors will use the QUOROM statement and generate ideas for its improvement.

Question: Therapy

Study Type: Meta-analysis

Learner Level: Intermediate

Notes: Important question which prompts good discussion of risks v. benefits v. cost. Good Discussion Points: Strengths: Well designed with clear description of methods. Can use methods to prompt vocabulary discussion for concepts pertaining to Meta-analysis (e.g. methods for validity assessment, summary treatment effects, sensitivity analysis, heterogeneity); Weaknesses: Variety of interventions can become very confusing to learners. More Advanced Discussion: might consider getting the QUOROM statement that is referenced in the methods of the paper (p. 859) for more in depth discussion of Meta-analysis methodology. Check list on p 1897 can be used to assess this (or any other systematic review); See also Hull 2001.

Moses-Kolko, E. L., et al. (2005). "Neonatal signs after late in utero exposure to serotonin reuptake inhibitors: literature review and implications for clinical applications." *JAMA* 293(19): 2372-2383.

CONTEXT: A neonatal behavioral syndrome linked to in utero serotonin reuptake inhibitor (SRI) exposure during the last trimester of pregnancy has been identified. The US Food and Drug Administration (FDA) and drug manufacturers have recently agreed to a class

labeling change for SRIs, which include selective serotonin reuptake inhibitors (SSRIs) and serotonin and norepinephrine reuptake inhibitors (SNRIs), to include information about potential adverse events in neonates exposed in utero. Integration of data about the neonatal behavioral syndrome into the management of pregnancy in women who take SRIs is a current challenge for physicians. OBJECTIVES: To review evidence regarding the SRI-related neonatal syndrome and to help clinicians guide their patients in a risk-benefit decision-making process. DATA SOURCES: We searched MEDLINE (1966-February 2005) and PsycINFO (1974-February 2005). All articles related to neonatal signs after in utero SRI exposure were acquired, as well as unpublished data on this topic from the FDA advisory committee meeting of June 2004. References cited in case reports and studies were reviewed. Foreign-language literature was included and translated to English. STUDY SELECTION AND DATA EXTRACTION: Studies were included if they had clearly identified maternal SRI exposure for a minimum of the final trimester of pregnancy through delivery and assessed neonatal outcomes. We identified 13 case reports describing a total of 18 cases. Nine cohort studies met criteria. When not included in the published article, relative risks and 95% confidence intervals (CIs) were computed from raw data and summary risk ratios and 95% CIs were determined with Mantel-Haenszel estimates. DATA SYNTHESIS: Compared with early gestational SRI exposure or no exposure, late SRI exposure carries an overall risk ratio of 3.0 (95% CI, 2.0-4.4) for a neonatal behavioral syndrome. The most SRI-related neonatal case reports involved fluoxetine and paroxetine exposures. Neonates primarily display central nervous system, motor, respiratory, and gastrointestinal signs that are usually mild and disappear by 2 weeks of age. Medical management has consisted primarily of supportive care in special care nurseries. A severe syndrome that consists of seizures, dehydration, excessive weight loss, hyperpyrexia, or intubation is rare in term infants (1/313 quantifiable cases). There have been no reported neonatal deaths attributable to neonatal SRI exposure. CONCLUSIONS: Available evidence indicates that in utero exposure to SRIs during the last trimester through delivery may result in a self-limited neonatal behavioral syndrome that can be managed with supportive care. The risks and benefits of discontinuing an SRI during pregnancy need to be carefully weighed for each individual patient. Development and validation of assessment methods and clinical management strategies are critical to advancing this research.

Question: Harm

Study Type: Systematic Review

Learner Level: Moderate to Advanced

Notes: Good article if you want to expose learners to systematic review addressing harm instead of therapy

Oler, A., et al. (1996). "Adding heparin to aspirin reduces the incidence of myocardial infarction and death in patients with unstable angina. A meta-analysis." *JAMA* 276(10): 811-815.

OBJECTIVE: To estimate the risk of myocardial infarction (MI) and death in patients with unstable angina who are treated with aspirin plus heparin compared with patients treated with aspirin alone. DATA SOURCES: Studies were retrieved using MEDLINE, bibliographies, and consultation with experts. STUDY SELECTION: Only published trials that enrolled patients with unstable angina, randomized participants to aspirin plus heparin vs aspirin alone, and reported incidence of myocardial infarction or death were included in the meta-analysis. DATA EXTRACTION: Patient outcomes including MI or death, recurrent ischemic pain, and major bleeding during randomized treatment; revascularization procedures after randomization; and MI or death during the 2 to 12 weeks following randomization were extracted by 2 authors, 1 of whom was blinded to the journal, institution, and author of each study. DATA SYNTHESIS: Six randomized trials were included. The overall summary relative risk (RR) of MI or death during randomized treatment was 0.67 (95% confidence interval [CI], 0.44-1.02) in patients with unstable angina treated with aspirin plus heparin compared with those treated with aspirin alone. The summary RRs for secondary endpoints in patients treated with aspirin plus heparin compared with those treated with aspirin alone were 0.68 (95% CI, 0.40-1.17) for recurrent ischemic pain; 0.82 (95% CI, 0.56-1.20) for MI or death 2 to 12 weeks following randomization; 1.03 (95% CI, 0.74-1.43) for revascularization; and 1.99 (95% CI, 0.52-7.65) for major bleeding. We found no statistically significant heterogeneity among individual study findings. CONCLUSIONS: Our findings are consistent with a 33% reduction in risk of MI or death in patients with unstable angina treated with aspirin plus heparin compared with those treated with aspirin alone. The bulk of evidence suggests that most patients with unstable angina should be treated with both heparin and aspirin.

Question: Therapy

Study Type: Systematic Review

Learner Level: Beginner / Intermediate

Notes: Important paper that directed clinical practice with respect to the use of aspirin and heparin. Good figures for discussing systematic review, confidence intervals, heterogeneity.

Pagliari, L., et al. (1992). "Prevention of first bleeding in cirrhosis. A meta-analysis of randomized trials of nonsurgical treatment." *Ann Intern Med* 117(1): 59-70.

OBJECTIVE: To assess the effectiveness of beta-blockers and endoscopic sclerotherapy in the prevention of first bleeding and reduction of mortality in patients with cirrhosis and esophagogastric varices. DATA SOURCES: Pertinent studies were selected using MEDLINE (1980 to 1990), reference lists from published articles or reviews, and congress abstract lists. STUDY SELECTION: Randomized trials comparing beta-blockers or sclerotherapy with a nonactive treatment. Nine randomized clinical trials of beta-blockers and 19 trials of sclerotherapy were reviewed. Seven trials of beta-blockers and 15 of sclerotherapy were published as full papers. DATA EXTRACTION: Crude rates of bleeding and death in treated and control groups were extracted from each trial by three independent observers according to the intention-to-treat principle. The quality of published papers was systematically assessed and scored. DATA SYNTHESIS: The Mantel-Haenszel-Peto method was used for statistical evaluation of heterogeneity and for pooling of

the results. No substantial heterogeneity was found, and the incidence of bleeding in trials of beta-blockers was significantly reduced (pooled odds ratio, 0.54; 95% CI, 0.39 to 0.74), particularly in patients with large or medium-sized varices or in those with varices and a hepatic vein pressure gradient above 12 mm Hg; however, only a trend toward reduced mortality was obtained. Sclerotherapy trials were highly heterogeneous in the direction of the treatment effects on both bleeding (pooled odds ratio, 0.6; CI, 0.49 to 0.74) and mortality (pooled odds ratio, 0.76; CI, 0.61 to 0.94). The quality of the trials and the rate of bleeding in the untreated groups were the major sources of heterogeneity. The favorable results of sclerotherapy were obtained in trials with high bleeding rates among controls; several of these trials had a low quality score. **CONCLUSIONS:** Beta-blockers may be recommended for prevention of first bleeding in cirrhotic patients with varices who have a high risk for bleeding. The effectiveness of sclerotherapy remains undetermined. Further trials in high-risk patients may prove useful if improved criteria to predict bleeding risk become available.

Question: Therapy

Study Type: Systematic Review

Learner Level: Intermediate / Advanced

Notes: Very good article for discussion of heterogeneity. (There is significant heterogeneity here...There are many good figures for illustration. However, the many figures and different kind of graphical representations may intimidate those who are not familiar with how to look at a systematic review. This is good for those who wish to take on a more challenging paper, but should be avoided by more novice meta-analysis learners.

Phung, O. J., et al. (2011). "Dosing frequency of unfractionated heparin thromboprophylaxis: a meta-analysis." *Chest* 140(2): 374-381.

BACKGROUND: In medical patients, it is unclear whether thromboprophylaxis with low-dose unfractionated heparin (UFH) should be administered bid or tid. **METHODS:** This study was a mixed-treatment comparison meta-analysis of randomized control trials that enrolled hospitalized nonsurgical patients at risk for VTE and compared UFH bid, UFH tid, or low-molecular-weight heparin (LMWH) to one another or to an inactive control subject. DVT, pulmonary embolism (PE), major bleeding, and death were measured. A Bayesian framework using a random-effects model was applied. **RESULTS:** Sixteen trials with moderate methodologic quality enrolling 27,667 patients contributed to this analysis. The relative risk and 95% credible intervals comparing UFH tid to UFH bid for DVT, PE, death, and major bleeding were 1.56 (0.64-4.33), 1.67 (0.49-208.09), 1.17 (0.72-1.95), and 0.89 (0.08-7.05), respectively. When compared with either dose of UFH, the use of LMWH has an effect similar to UFH on all four outcomes. **CONCLUSIONS:** Moderate-quality evidence suggests that subcutaneous UFH bid and UFH tid do not differ in effect on DVT, PE, major bleeding, and mortality. Either of the two dosing regimens of UFH or LMWH appears to be a reasonable strategy for thromboprophylaxis in medical patients. A future randomized trial comparing the two doses of UFH is very unlikely, considering the very large sample size that would be required to demonstrate a significant difference, which, if it exists, is undoubtedly small.

Question: Therapy

Study Type: Meta-analysis

Learner Level: Beginner

Notes: .

Pimpin, L., et al. (2016). "Is Butter Back? A Systematic Review and Meta-Analysis of Butter Consumption and Risk of Cardiovascular Disease, Diabetes, and Total Mortality." *PLoS One* 11(6): e0158118.

BACKGROUND: Dietary guidelines recommend avoiding foods high in saturated fat. Yet, emerging evidence suggests cardiometabolic benefits of dairy products and dairy fat. Evidence on the role of butter, with high saturated dairy fat content, for total mortality, cardiovascular disease, and type 2 diabetes remains unclear. We aimed to systematically review and meta-analyze the association of butter consumption with all-cause mortality, cardiovascular disease, and diabetes in general populations. **METHODS AND FINDINGS:** We searched 9 databases from inception to May 2015 without restriction on setting, or language, using keywords related to butter consumption and cardiometabolic outcomes. Prospective cohorts or randomized clinical trials providing estimates of effects of butter intake on mortality, cardiovascular disease including coronary heart disease and stroke, or diabetes in adult populations were included. One investigator screened titles and abstracts; and two reviewed full-text articles independently in duplicate, and extracted study and participant characteristics, exposure and outcome definitions and assessment methods, analysis methods, and adjusted effects and associated uncertainty, all independently in duplicate. Study quality was evaluated by a modified Newcastle-Ottawa score. Random and fixed effects meta-analysis pooled findings, with heterogeneity assessed using the I² statistic and publication bias by Egger's test and visual inspection of funnel plots. We identified 9 publications including 15 country-specific cohorts, together reporting on 636,151 unique participants with 6.5 million person-years of follow-up and including 28,271 total deaths, 9,783 cases of incident cardiovascular disease, and 23,954 cases of incident diabetes. No RCTs were identified. Butter consumption was weakly associated with all-cause mortality (N = 9 country-specific cohorts; per 14g(1 tablespoon)/day: RR = 1.01, 95%CI = 1.00, 1.03, P = 0.045); was not significantly associated with any cardiovascular disease (N = 4; RR = 1.00, 95%CI = 0.98, 1.02; P = 0.704), coronary heart disease (N = 3; RR = 0.99, 95%CI = 0.96, 1.03; P = 0.537), or stroke (N = 3; RR = 1.01, 95%CI = 0.98, 1.03; P = 0.737), and was inversely associated with incidence of diabetes (N = 11; RR = 0.96, 95%CI = 0.93, 0.99; P = 0.021). We did not identify evidence for heterogeneity nor publication bias. **CONCLUSIONS:** This systematic review and meta-analysis suggests relatively small or neutral overall associations of butter with mortality, CVD, and diabetes. These findings do not support a need for major emphasis in dietary guidelines on either increasing or decreasing butter consumption, in comparison to other better established dietary priorities; while also highlighting the need for additional investigation of health and metabolic effects of butter and dairy fat.

Question: Harm

Study Type: Systematic Review

Learner Level: Beginner

Notes: Good for teaching Forest Plots and Heterogeneity; hits all the critical appraisal validity criteria including funnel plots.

Rathbun, S. W., et al. (2000). "Sensitivity and specificity of helical computed tomography in the diagnosis of pulmonary embolism: a systematic review." *Ann Intern Med* 132(3): 227-232.

PURPOSE: To determine the sensitivity and specificity of helical computed tomography (CT) for the diagnosis of pulmonary embolism and to determine the safety of withholding anticoagulant therapy in patients who have clinically suspected pulmonary embolism and negative results on helical CT. **DATA SOURCES:** The MEDLINE database was searched for all reports published from 1986 to October 1999 that evaluated the use of helical CT for the diagnosis of pulmonary embolism. Bibliographies of the retrieved articles were cross-checked to identify additional studies. **STUDY SELECTION:** All prospective English-language studies were selected. Retrospective studies, review articles, and case reports were excluded, and 5 of the 20 identified articles were excluded. The scientific validity of the remaining 15 articles was assessed. **DATA EXTRACTION:** Two of the authors used a priori, pre-defined criteria to independently assess each study. A third author resolved disagreements by adjudication. The pre-defined criteria were inclusion of a consecutive series of all patients with suspected pulmonary embolism, inclusion of patients with and those without pulmonary embolism, a broad spectrum of patient characteristics, performance of helical CT and pulmonary angiography (or an appropriate reference test) in all patients, and independent interpretation of the CT scan and pulmonary angiogram (or reference test). Specific data on sensitivity and specificity and the associated 95% CIs were recorded when available. **DATA SYNTHESIS:** No study met all of the predefined criteria for adequately evaluating sensitivity and specificity. The reported sensitivity of helical CT ranged from 53% to 100%, and specificity ranged from 81% to 100%. In no prospective study was anticoagulant therapy withheld without further testing for venous thromboembolism in consecutive patients with suspected pulmonary embolism. One prospective study reported the outcome of selected patients with negative results on helical CT who did not receive anticoagulant therapy. **CONCLUSIONS:** Use of helical CT in the diagnosis of pulmonary embolism has not been adequately evaluated. The safety of withholding anticoagulant treatment in patients with negative results on helical CT is uncertain. Definitive large, prospective studies should be done to evaluate the sensitivity, specificity, and safety of helical CT for diagnosis of suspected pulmonary embolism.

Question: Diagnosis

Study Type: Systematic Review

Learner Level: Intermediate / Advanced

Notes: Important paper that brought together the evidence regarding helical CT. Good discussion points: Good for discussion of how papers are graded in a systematic review. Note: this paper predates the other diagnostic paper about helical CT in the diagnosis section. Interesting to consider whether the two papers come to similar conclusions. See also Perrier A. *Annals of Internal Medicine* 2001;125(2):88-97 for example of an original study

Ray, K. K., et al. (2010). "Statins and all-cause mortality in high-risk primary prevention: a meta-analysis of 11 randomized controlled trials involving 65,229 participants." *Arch Intern Med* 170(12): 1024-1031.

BACKGROUND: Statins have been shown to reduce the risk of all-cause mortality among individuals with clinical history of coronary heart disease. However, it remains uncertain whether statins have similar mortality benefit in a high-risk primary prevention setting. Notably, all systematic reviews to date included trials that in part incorporated participants with prior cardiovascular disease (CVD) at baseline. Our objective was to reliably determine if statin therapy reduces all-cause mortality among intermediate to high-risk individuals without a history of CVD. **DATA SOURCES:** Trials were identified through computerized literature searches of MEDLINE and Cochrane databases (January 1970-May 2009) using terms related to statins, clinical trials, and cardiovascular end points and through bibliographies of retrieved studies. **STUDY SELECTION:** Prospective, randomized controlled trials of statin therapy performed in individuals free from CVD at baseline and that reported details, or could supply data, on all-cause mortality. **DATA EXTRACTION:** Relevant data including the number of patients randomized, mean duration of follow-up, and the number of incident deaths were obtained from the principal publication or by correspondence with the investigators. **DATA SYNTHESIS:** Data were combined from 11 studies and effect estimates were pooled using a random-effects model meta-analysis, with heterogeneity assessed with the I^2 statistic. Data were available on 65,229 participants followed for approximately 244,000 person-years, during which 2793 deaths occurred. The use of statins in this high-risk primary prevention setting was not associated with a statistically significant reduction (risk ratio, 0.91; 95% confidence interval, 0.83-1.01) in the risk of all-cause mortality. There was no statistical evidence of heterogeneity among studies ($I^2 = 23\%$; 95% confidence interval, 0%-61% [$P = .23$]). **CONCLUSION:** This literature-based meta-analysis did not find evidence for the benefit of statin therapy on all-cause mortality in a high-risk primary prevention set-up.

Question: Therapy

Study Type: Meta-analysis

Learner Level: Beginner

Notes: Well done large meta-analysis of the use of statins for primary prevention in high risk patients without cardiac events. Beginner for a meta-analysis in that it has a manageable amount of data presented, forest plots are straightforward with reporting of I-squared

Rodrigo, G. J., et al. (2017). "LABA/LAMA combinations versus LAMA monotherapy or LABA/ICS in COPD: a systematic review and meta-analysis." *International Journal of Chronic Obstructive Pulmonary Disease* 12: 907-922.

BACKGROUND: Randomized controlled trials (RCTs) indicate that long-acting bronchodilator combinations, such as beta2-agonist (LABA)/muscarinic antagonist (LAMA), have favorable efficacy compared with commonly used COPD treatments. The objective of this analysis was to compare the efficacy and safety of LABA/LAMA with LAMA or LABA/inhaled corticosteroid (ICS) in adults with stable moderate-to-very-severe COPD. **METHODS:** This systematic review and meta-analysis (PubMed/MEDLINE, Embase, Cochrane Library and clinical trial/manufacture databases) included RCTs comparing ≥ 12 weeks' LABA/LAMA treatment with LAMA and/or LABA/ICS (approved doses only). Eligible studies were independently selected by two authors using predefined data fields; the Preferred Reporting Items for Systematic Reviews and Meta-Analyses guidelines were followed. **RESULTS:** Eighteen studies (23 trials) were eligible (N=20,185). LABA/LAMA significantly improved trough forced expiratory volume in 1 second (FEV1) from baseline to week 12 versus both LAMA and LABA/ICS (0.07 L and 0.08 L, $P < 0.0001$), with patients more likely to achieve clinically important improvements in FEV1 of > 100 mL (risk ratio [RR]: 1.33, 95% confidence interval [CI]: [1.20, 1.46] and RR: 1.44, 95% CI: [1.33, 1.56], respectively, the number needed to treat being eight and six, respectively). LABA/LAMA improved transitional dyspnea index and St George's Respiratory Questionnaire scores at week 12 versus LAMA (both $P < 0.0001$), but not versus LABA/ICS, and reduced rescue medication use versus both ($P < 0.0001$ and $P = 0.001$, respectively). LABA/LAMA significantly reduced moderate/severe exacerbation rate compared with LABA/ICS (RR 0.82, 95% CI: [0.75, 0.91]). Adverse event (AE) incidence was no different for LABA/LAMA versus LAMA treatment, but it was lower versus LABA/ICS (RR 0.94, 95% CI: [0.89, 0.99]), including a lower pneumonia risk (RR 0.59, 95% CI: [0.43, 0.81]). LABA/LAMA presented a lower risk for withdrawals due to lack of efficacy versus LAMA (RR: 0.66, 95% CI: [0.51, 0.87]) and due to AEs versus LABA/ICS (RR: 0.83, 95% CI: [0.69, 0.99]). **CONCLUSION:** The greater efficacy and comparable safety profiles observed with LABA/LAMA combinations versus LAMA or LABA/ICS support their potential role as first-line treatment options in COPD. These findings are of direct relevance to clinical practice because we included all currently available LABA/LAMAs and comparators, only at doses approved for clinical use.

Question: Therapy

Study Type: Systematic Review / Meta-Analysis

Learner Level: intermediate

Notes: Good to discuss heterogeneity.

Ross, J. R., et al. (2003). "Systematic review of role of bisphosphonates on skeletal morbidity in metastatic cancer." *BMJ* 327(7413): 469.

OBJECTIVE: To review the evidence for the use of bisphosphonates to reduce skeletal morbidity in cancer patients with bone metastases. **DATA SOURCES:** Electronic databases, scanning reference lists, and consultation with experts and pharmaceutical companies. Foreign language papers were included. **STUDY SELECTION:** Included trials were randomised controlled trials of patients with malignant disease and bone metastases who were treated with oral or intravenous bisphosphonate compared with another bisphosphonate, placebo, or standard care. All trials measured at least one outcome of skeletal morbidity. **RESULTS:** 95 articles were identified; 30 studies fulfilled inclusion criteria. In studies that lasted ≥ 6 months, compared with placebo bisphosphonates significantly reduced the odds ratio for fractures (vertebral 0.69, 95% confidence interval 0.57 to 0.84, $P < 0.0001$; non-vertebral 0.65, 0.54 to 0.79, $P < 0.0001$; combined 0.65, 0.55 to 0.78, $P < 0.0001$), radiotherapy (0.67, 0.57 to 0.79, $P < 0.0001$), and hypercalcaemia (0.54, 0.36 to 0.81, $P = 0.003$) but not for orthopaedic surgery (0.70, 0.46 to 1.05, $P = 0.086$) or spinal cord compression (0.71, 0.47 to 1.08, $P = 0.113$). The reduction in orthopaedic surgery was significant in studies that lasted over a year (0.59, 0.39 to 0.88, $P = 0.009$). Use of bisphosphonates significantly increased time to first skeletal related event but did not increase survival. Subanalyses showed that most evidence supports use of intravenous aminobisphosphonates. **CONCLUSIONS:** In people with metastatic bone disease bisphosphonates significantly decrease skeletal morbidity, except for spinal cord compression and increased time to first skeletal related event. Treatment should start when bone metastases are diagnosed and continue until it is no longer clinically relevant.

Question: Therapy

Study Type: Meta-analysis

Learner Level: Intermediate / Advanced

Notes: Strong methodology Meta-analysis with very clear description of methods (Figure 1 nice overview for flow of selected papers with on-line access to even more complete details). Good Discussion Points: Lots of great forest plots! / Both endpoints that use straight RR and those needing conversion to effect size / No data table in paper, but available from a website. Generates great discussion about homogeneity and generalizability (includes lytic cancers, blastic cancers, wide population and age range). Good paper to point out pitfall of trying to get a systematic review to answer more than its focused clinical question (e.g., which bisphosphonate should you try first?)

Rothman, R., et al. (2003). "Does this child have acute otitis media?" *JAMA* 290(12): 1633-1640.

CONTEXT: Acute otitis media (AOM) is one of the most common problems in pediatrics. An accurate diagnosis of AOM can guide proper treatment and follow-up. OBJECTIVE: To systematically review the literature regarding precision and accuracy of history taking and physical examination in diagnosing AOM in children. DATA SOURCES: We searched MEDLINE for English-language articles published from 1966 through May 2002. Bibliographies of retrieved articles and textbooks were also searched. STUDY SELECTION: We located studies with original data on the precision or accuracy of history or physical examination for AOM in children. Of 397 references initially identified, 6 met inclusion criteria for analysis. DATA EXTRACTION: Two authors independently reviewed and abstracted data to calculate likelihood ratios (LRs) for symptoms and signs. DATA SYNTHESIS: Four studies of symptoms used clinical diagnosis as the criterion standard and were limited by incorporation bias. Ear pain is the most useful symptom (positive LRs, 3.0-7.3); fever, upper respiratory tract symptoms, and irritability are less useful. One study of clinical signs used tympanocentesis as the criterion standard, and we adjusted the results to correct for verification bias. A cloudy (adjusted LR, 34; 95% confidence interval [CI], 28-42), bulging (adjusted LR, 51; 95% CI, 36-73), or distinctly immobile (adjusted LR, 31; 95% CI, 26-37) tympanic membrane on pneumatic otoscopy are the most useful signs for detecting AOM. A distinctly red tympanic membrane is also helpful (adjusted LR, 8.4; 95% CI, 6.7-11) whereas a normal color makes AOM much less likely (adjusted LR, 0.2; 95% CI, 0.19-0.21). CONCLUSIONS: Although many of the studies included in this analysis are limited by bias, a cloudy, bulging, or clearly immobile tympanic membrane is most helpful for detecting AOM. The degree of erythema may also be useful since a normal color makes otitis media unlikely whereas a distinctly red tympanic membrane increases the likelihood significantly.

Question: Diagnosis

Study Type: Meta-analysis

Learner Level: Intermediate / Advanced

Notes: Another meta-analysis is part of the Rational Clinical Exam Series. This article on otitis media in children is relevant to anyone who has ever had (or ever will have) a cranky child with a fever...Good discussion points: As with all Rational Clinical Exam articles, this is an evidence summary of diagnostic tests- thus one can discuss both systematic review methodology AND diagnosis, specifically Likelihood ratio. Excellent paragraph under statistical methods (Page 990) that defines likelihood ratio as well as diagnostic odds ratio. Stumbling block may be the large number of items listed in Tables 2 and 3 (p 992 and 993), however a clear difference in data can be noted in patients older than 60 years compared to all comers.

Salpeter, S. R., et al. (2004). "Cardiovascular effects of beta-agonists in patients with asthma and COPD: a meta-analysis." *Chest* 125(6): 2309-2321.

BACKGROUND: beta-Adrenergic agonists exert physiologic effects that are the opposite of those of beta-blockers. beta-Blockers are known to reduce morbidity and mortality in patients with cardiac disease. beta(2)-Agonist use in patients with obstructive airway disease has been associated with an increased risk for myocardial infarction, congestive heart failure, cardiac arrest, and acute cardiac death. OBJECTIVES: To assess the cardiovascular safety of beta(2)-agonist use in patients with obstructive airway disease, defined as asthma or COPD. METHODS: A meta-analysis of randomized placebo-controlled trials of beta(2)-agonist treatment in patients with obstructive airway disease was performed, to evaluate the short-term effect on heart rate and potassium concentrations, and the long-term effect on adverse cardiovascular events. Longer duration trials were included in the analysis if they reported at least one adverse event. Adverse events included sinus and ventricular tachycardia, syncope, atrial fibrillation, congestive heart failure, myocardial infarction, cardiac arrest, or sudden death. RESULTS: Thirteen single-dose trials and 20 longer duration trials were included in the study. A single dose of beta(2)-agonist increased the heart rate by 9.12 beats/min (95% confidence interval [CI], 5.32 to 12.92) and reduced the potassium concentration by 0.36 mmol/L (95% CI, 0.18 to 0.54), compared to placebo. For trials lasting from 3 days to 1 year, beta(2)-agonist treatment significantly increased the risk for a cardiovascular event (relative risk [RR], 2.54; 95% CI, 1.59 to 4.05) compared to placebo. The RR for sinus tachycardia alone was 3.06 (95% CI, 1.70 to 5.50), and for all other events it was 1.66 (95% CI, 0.76 to 3.6). CONCLUSION: beta(2)-Agonist use in patients with obstructive airway disease increases the risk for adverse cardiovascular events. The initiation of treatment increases heart rate and reduces potassium concentrations compared to placebo. It could be through these mechanisms, and other effects of beta-adrenergic stimulation, that beta(2)-agonists may precipitate ischemia, congestive heart failure, arrhythmias, and sudden death.

Question: Therapy

Study Type: Meta-analysis

Learner Level: Beginner

Notes: Good to teach surrogate outcomes

Salpeter, S. R., et al. (2010). "Long-acting beta-agonists with and without inhaled corticosteroids and catastrophic asthma events." *Am J Med* 123(4): 322-328 e322.

BACKGROUND: It is unclear whether long-acting beta-agonists with concomitant inhaled corticosteroids increase asthma-related intubations and deaths. We pooled data on long-acting beta-agonists with variable and concomitant inhaled corticosteroids to evaluate the risk for catastrophic asthma events. METHODS: We conducted searches of electronic databases, the US Food and Drug Administration website, clinical-trials registries, and selected references through December 2008. We analyzed randomized controlled trials in patients with asthma, which lasted at least 3 months, evaluated long-acting beta-agonists compared with placebo or long-acting beta-agonists with inhaled corticosteroids compared with corticosteroids alone, and included at least 1 catastrophic event, defined as asthma-related intubation or death. RESULTS: In pooled trial data that included 36,588 participants, long-acting beta-agonists increased catastrophic events 2-fold (Peto odds ratio [OR] 2.10; 95% confidence interval [CI], 1.37-3.22). Statistically

significant increases were seen for long-acting beta-agonists with variable corticosteroids compared with placebo (OR 1.83; 95% CI, 1.14-2.95) and for concomitant treatment with corticosteroids compared with corticosteroids alone (OR 3.65; 95% CI, 1.39-9.55). Similar increases in risk were seen for variable and concomitant corticosteroid use, salmeterol and formoterol, and children and adults. When the analysis was restricted to trials with controlled corticosteroid use, given as part of the study intervention, concomitant treatment still increased catastrophic events compared with corticosteroids alone (OR 8.19; 95% CI, 1.10-61.18). CONCLUSION: Long-acting beta-agonists increase the risk for asthma-related intubations and deaths, even when used in a controlled fashion with concomitant inhaled corticosteroids.

Question: Therapy

Study Type: Meta-analysis

Learner Level: Beginner - Intermediate

Notes: controversial topic; RCTs are summarized with outcome of increased risk (harm as opposed to the usual framework of looking for benefit); Good to discuss ethics (the RCTs were done to show benefit, but didn't end up showing that in all cases...); Nice Forest plots including use of I-squared; Article accompanied by an ACP-JC

Santangeli, P., et al. (2010). "Meta-analysis: age and effectiveness of prophylactic implantable cardioverter-defibrillators." Ann Intern Med 153(9): 592-599.

BACKGROUND: Implantable cardioverter-defibrillators (ICDs) for the primary prevention of sudden cardiac death have been proven effective in several clinical trials. PURPOSE: To summarize evidence about the effectiveness of ICDs versus standard medical therapy for the primary prevention of sudden cardiac death in different age groups of patients with severe left ventricular dysfunction. DATA SOURCES: MEDLINE, Embase, CENTRAL, BioMed Central, Cardiosource, ClinicalTrials.gov, and ISI Web of Science (January 1970 to April 2010) were searched with no language restrictions. STUDY SELECTION: Two independent reviewers screened titles and abstracts to identify randomized, controlled trials of prophylactic ICD versus medical therapy in patients with severe left ventricular dysfunction that provided data about mortality outcomes for different age groups. DATA EXTRACTION: Two independent reviewers assessed risk for bias of trials and extracted patient and study characteristics and hazard ratios (HRs) relevant to all-cause mortality. DATA SYNTHESIS: Five trials (MADIT-II, DEFINITE, DINAMIT, SCD-HeFT, and IRIS) that enrolled 5783 patients (44% were elderly) were included. The primary analysis, which excluded the 2 trials enrolling patients early after acute myocardial infarction (DINAMIT and IRIS), found that prophylactic ICD therapy reduced mortality in younger patients (HR, 0.65 [95% CI, 0.50 to 0.83]; $P < 0.001$). A smaller survival benefit was found in elderly patients (HR, 0.75 [95% CI, 0.61 to 0.91]) that was not confirmed when MADIT-II patients older than 70 years were excluded or when data from DINAMIT and IRIS were included [corrected]. LIMITATIONS: Four potentially eligible trials were not included in the meta-analysis because mortality data by age group were not available. Adjustment for differences in comorbid conditions and medical therapies among patients enrolled in the trials was not possible. CONCLUSION: Available data suggest that prophylactic ICD therapy may be less beneficial for elderly patients with severe left ventricular dysfunction than for younger patients [corrected]. PRIMARY FUNDING SOURCE: None.

Question: Therapy

Study Type: Meta-analysis

Learner Level: Intermediate

Notes: Nice example of a systematic review in examining subgroup efficacy. Good clear assessment of evidence quality, forest plots, discussion points on heterogeneity. Moderate difficulty.

Shekelle, P. G., et al. (2004). "Effect of supplemental vitamin E for the prevention and treatment of cardiovascular disease." Journal of General Internal Medicine 19(4): 380-389.

OBJECTIVE: To evaluate and synthesize the evidence on the effect of supplements of vitamin E on the prevention and treatment of cardiovascular disease. DESIGN: Systematic review of placebo-controlled randomized controlled trials; meta-analysis where justified. MEASUREMENTS AND MAIN RESULTS: Eighty-four eligible trials were identified. For the outcomes of all-cause mortality, cardiovascular mortality, fatal or nonfatal myocardial infarction, and blood lipids, neither supplements of vitamin E alone nor vitamin E given with other agents yielded a statistically significant beneficial or adverse pooled relative risk (for example, pooled relative risk of vitamin E alone = 0.96 [95% confidence interval (CI), 0.84 to 1.10]; 0.97 [95% CI, 0.80 to 1.90]; and 0.72 [95% CI, 0.51 to 1.02] for all-cause mortality, cardiovascular mortality, and nonfatal myocardial infarction, respectively. CONCLUSIONS: There is good evidence that vitamin E supplementation does not beneficially or adversely affect cardiovascular outcomes.

Question: Therapy

Study Type: Systematic Review

Learner Level: Beginner

Notes: High quality systematic review. Easy to understand therapy and outcomes if teaching a group from mixed specialties. It is a treatment that many people are still on contrary to the evidence. A little bit complicated since it looked at vitamin E alone and vitamin E in combination as 2 separate analyses in the same paper. Good Discussion Points: Validity of SR since it is extremely well done with extensive search and description of methodology. Discussion of heterogeneity. Main outcome did not show

heterogeneity, but the secondary outcomes did. This can be seen statistically as well as in the Forest plots. They also mention the I² statistic as well as χ^2 , so can use this for advanced learners. Given the differences between studies it is a good example of when to pool and when not to. I focus on the main outcome of mortality to try and simplify this long article. ACP Journal Club reviews this article. See also Miller, E. *Annals of Internal Medicine* Jan 4 2005 and Eidelman RS. *Annals of Internal Medicine* 2004.

Sirven, J. I., et al. (2004). "Seizure prophylaxis in patients with brain tumors: a meta-analysis." *Mayo Clin Proc* 79(12): 1489-1494.

OBJECTIVE: To assess whether antiepileptic drugs (AEDs) should be prescribed to patients with brain tumors who have no history of seizures. **METHODS:** We performed a meta-analysis of randomized controlled trials (1966-2004) that evaluated the efficacy of AED prophylaxis vs no treatment or placebo to prevent seizures in patients with brain tumors who had no history of epilepsy. Summary odds ratios (ORs) were calculated using a random-effects model. Three subanalyses were performed to assess pooled ORs of seizures in patients with primary glial tumors, cerebral metastases, and meningiomas. **RESULTS:** Of 474 articles found in the initial search, 17 were identified as primary studies. Five trials met inclusion criteria: patients with a neoplasm (primary glial tumors, cerebral metastases, and meningiomas) but no history of epilepsy who were randomized to either an AED or placebo. The 3 AEDs studied were phenobarbital, phenytoin, and valproic acid. Of the 5 trials, 4 showed no statistical benefit of seizure prophylaxis with an AED. Meta-analysis confirmed the lack of AED benefit at 1 week (OR, 0.91; 95% confidence interval [CI], 0.45-1.83) and at 6 months (OR, 1.01; 95% CI, 0.51-1.98) of follow-up. The AEDs had no effect on seizure prevention for specific tumor pathology, including primary glial tumors (OR, 3.46; 95% CI, 0.32-37.47), cerebral metastases (OR, 2.50; 95% CI, 0.25-24.72), and meningiomas (OR, 0.62; 95% CI, 0.10-3.85). **CONCLUSIONS:** No evidence supports AED prophylaxis with phenobarbital, phenytoin, or valproic acid in patients with brain tumors and no history of seizures, regardless of neoplastic type. Subspecialists who treat patients with brain tumors need more education on this issue. Future randomized controlled trials should address whether any of the newer AEDs are useful for seizure prophylaxis.

Question: Therapy

Study Type: Meta-analysis

Learner Level: Intermediate

Notes: Good to teach creating your own forest plot

Stern, A., et al. (2017). "Corticosteroids for pneumonia." *Cochrane Database Syst Rev* 12: Cd007720.

BACKGROUND: Pneumonia is a common and potentially serious illness. Corticosteroids have been suggested for the treatment of different types of infection, however their role in the treatment of pneumonia remains unclear. This is an update of a review published in 2011. **OBJECTIVES:** To assess the efficacy and safety of corticosteroids in the treatment of pneumonia. **SEARCH METHODS:** We searched the Cochrane Acute Respiratory Infections Group's Specialised Register, CENTRAL, MEDLINE, Embase, and LILACS on 3 March 2017, together with relevant conference proceedings and references of identified trials. We also searched three trials registers for ongoing and unpublished trials. **SELECTION CRITERIA:** We included randomised controlled trials (RCTs) that assessed systemic corticosteroid therapy, given as adjunct to antibiotic treatment, versus placebo or no corticosteroids for adults and children with pneumonia. **DATA COLLECTION AND ANALYSIS:** We used standard methodological procedures expected by Cochrane. Two review authors independently assessed risk of bias and extracted data. We contacted study authors for additional information. We estimated risk ratios (RR) with 95% confidence intervals (CI) and pooled data using the Mantel-Haenszel fixed-effect model when possible. **MAIN RESULTS:** We included 17 RCTs comprising a total of 2264 participants; 13 RCTs included 1954 adult participants, and four RCTs included 310 children. This update included 12 new studies, excluded one previously included study, and excluded five new trials. One trial awaits classification. All trials limited inclusion to inpatients with community-acquired pneumonia (CAP), with or without healthcare-associated pneumonia (HCAP). We assessed the risk of selection bias and attrition bias as low or unclear overall. We assessed performance bias risk as low for nine trials, unclear for one trial, and high for seven trials. We assessed reporting bias risk as low for three trials and high for the remaining 14 trials. Corticosteroids significantly reduced mortality in adults with severe pneumonia (RR 0.58, 95% CI 0.40 to 0.84; moderate-quality evidence), but not in adults with non-severe pneumonia (RR 0.95, 95% CI 0.45 to 2.00). Early clinical failure rates (defined as death from any cause, radiographic progression, or clinical instability at day 5 to 8) were significantly reduced with corticosteroids in people with severe and non-severe pneumonia (RR 0.32, 95% CI 0.15 to 0.7; and RR 0.68, 95% CI 0.56 to 0.83, respectively; high-quality evidence). Corticosteroids reduced time to clinical cure, length of hospital and intensive care unit stays, development of respiratory failure or shock not present at pneumonia onset, and rates of pneumonia complications. Among children with bacterial pneumonia, corticosteroids reduced early clinical failure rates (defined as for adults, RR 0.41, 95% CI 0.24 to 0.70; high-quality evidence) based on two small, clinically heterogeneous trials, and reduced time to clinical cure. Hyperglycaemia was significantly more common in adults treated with corticosteroids (RR 1.72, 95% CI 1.38 to 2.14). There were no significant differences between corticosteroid-treated people and controls for other adverse events or secondary infections (RR 1.19, 95% CI 0.73 to 1.93). **AUTHORS' CONCLUSIONS:** Corticosteroid therapy reduced mortality and morbidity in adults with severe CAP; the number needed to treat for an additional beneficial outcome was 18 patients (95% CI 12 to 49) to prevent one death. Corticosteroid therapy reduced morbidity, but not mortality, for adults and children with non-severe CAP. Corticosteroid therapy was associated with more adverse events, especially hyperglycaemia, but the harms did not seem to outweigh the benefits.

Question: Therapy

Study Type: Systematic Review

Learner Level: Intermediate

Notes: Teach this paper in conjunction with the Clinical Infectious Disease systematic review by Briel et al. in order to demonstrate that systematic reviews done by different authorities may have conflicting results and that it is important to understand the methods the authors use to arrive at their conclusions.

Stevens, B., et al. (1997). "The efficacy of sucrose for relieving procedural pain in neonates--a systematic review and meta-analysis." *Acta Paediatrica* 86(8): 837-842.

The objective was to determine the efficacy and optimal dose of sucrose for relieving procedural pain in neonates. Data were obtained using MEDLINE, EMBASE, Reference Update and personal files and assessed for quality of the methods. Data from all randomized controlled trials where term and preterm neonates received a heelstick or venipuncture were examined for the efficacy of different sucrose doses (0.18 g, 0.24 g, 0.48 g or 0.50 g, 1.0 g) and water (placebo). The primary outcome was the proportion of time crying during 3 min after the painful stimulus. Data were combined across studies using a random effects model, adapted for use with single groups, producing a point estimate and 95% confidence interval (CI). Thirteen trials were identified; eight were rejected as data were inappropriate, non-extractable, or the primary outcome was not measured. Five studies provided data on 271 infants. The proportion of time crying did not differ between 0.18 g of sucrose and water ($p > 0.05$) but was significantly lower in all other sucrose groups. There were no differences in proportion of time crying between term and preterm neonates. Sucrose reduced the proportion of time crying during painful procedures in neonates. The 0.18 g dose of sucrose was ineffective. Doses of 0.24 g (2 ml of 12% sucrose solution) were most effective. A dose of 0.50 g provided no additional benefit.

Question: Therapy

Study Type: Systematic Review

Learner Level: Beginner / Intermediate

Notes: Straightforward methods that describe both the selection of articles as well as the assessment of quality of the methods. Good discussion points: Good paper for anyone who has ever had a child that underwent a procedure.

Thangaratinam, S., et al. (2012). "Pulse oximetry screening for critical congenital heart defects in asymptomatic newborn babies: a systematic review and meta-analysis." *Lancet* 379(9835): 2459-2464.

BACKGROUND: Screening for critical congenital heart defects in newborn babies can aid in early recognition, with the prospect of improved outcome. We assessed the performance of pulse oximetry as a screening method for the detection of critical congenital heart defects in asymptomatic newborn babies. **METHODS:** In this systematic review, we searched Medline (1951-2011), Embase (1974-2011), Cochrane Library (2011), and Scisearch (1974-2011) for relevant citations with no language restriction. We selected studies that assessed the accuracy of pulse oximetry for the detection of critical congenital heart defects in asymptomatic newborn babies. Two reviewers selected studies that met the predefined criteria for population, tests, and outcomes. We calculated sensitivity, specificity, and corresponding 95% CIs for individual studies. A hierarchical receiver operating characteristic curve was fitted to generate summary estimates of sensitivity and specificity with a random effects model. **FINDINGS:** We screened 552 studies and identified 13 eligible studies with data for 229,421 newborn babies. The overall sensitivity of pulse oximetry for detection of critical congenital heart defects was 76.5% (95% CI 67.7-83.5). The specificity was 99.9% (99.7-99.9), with a false-positive rate of 0.14% (0.06-0.33). The false-positive rate for detection of critical congenital heart defects was particularly low when newborn pulse oximetry was done after 24 h from birth than when it was done before 24 h (0.05% [0.02-0.12] vs 0.50 [0.29-0.86]; $p=0.0017$). **INTERPRETATION:** Pulse oximetry is highly specific for detection of critical congenital heart defects with moderate sensitivity, that meets criteria for universal screening. **FUNDING:** None.

Question: Diagnosis

Study Type: Systematic Review

Learner Level: Advanced

Notes: Includes data from the Granelli study. Overall well-done, though no traditional forest plots. Good for teaching heterogeneity secondary to differences in application of test, displayed in a format similar to forest plots (24 hours post birth, foot, hand etc.). Also has a funnel plot demonstrating possible publication bias.

Tuon, F. F., et al. (2006). "Adenosine deaminase and tuberculous pericarditis--a systematic review with meta-analysis." *Acta Trop* 99(1): 67-74.

BACKGROUNDS: Adenosine deaminase (ADA) activity in pericardial fluid is a valuable aid in the diagnosis of tuberculous pericarditis (TP), but there is no systematic review performed to evaluate the benefits of ADA activity as an adjunctive test for TP diagnosis. The objective of this systematic review was to evaluate the utility of ADA activity as a diagnostic marker of TP on patients presenting with pericardial effusion. **METHODS:** MEDLINE, LILACS and Cochrane Library databases (1980-2005) searches to identify articles related to adenosine deaminase activity on TP diagnosis. Articles with patients with at least one TP diagnostic criteria were included. The controls were patients with other pericardial diseases with moderate or large pericardial effusion. To calculate the sensitivity, specificity, as well as positive and negative likelihood ratios we extracted the total number of confirmed TP cases over all patients with pericardial effusion as well as the number of cases with ADA activity values of 40 U/L and over. **RESULTS:** Thirty one studies met our initial inclusion criteria and five articles were selected. The heterogeneity limited the specificity analysis ($p=0.004$). The method yielded a sensitivity and specificity of 88% and 83%, respectively. The SROC curve presented an area with a tendency towards 1

(value of 0.9539) and corroborates the diagnostic value of ADA activity. CONCLUSIONS: The present study confirms the clinical value of ADA activity as adjunctive diagnostic marker of TP among other causes of pericardial effusion.

Question: Diagnosis

Study Type: Meta-analysis

Learner Level: Advanced

Notes: Good to teach likelihood ratios

Turpie, A. G., et al. (2002). "Fondaparinux vs enoxaparin for the prevention of venous thromboembolism in major orthopedic surgery: a meta-analysis of 4 randomized double-blind studies." Arch Intern Med 162(16): 1833-1840.

BACKGROUND: Orthopedic surgery remains a condition at high risk of venous thromboembolism (VTE). Fondaparinux, the first of a new class of synthetic selective factor Xa inhibitors, may further reduce this risk compared with currently available thromboprophylactic treatments. METHODS: A meta-analysis of 4 multicenter, randomized, double-blind trials in patients undergoing elective hip replacement, elective major knee surgery, and surgery for hip fracture (N = 7344) was performed to determine whether a subcutaneous 2.5-mg, once-daily regimen of fondaparinux sodium starting 6 hours after surgery was more effective and as safe as approved enoxaparin regimens in preventing VTE. The primary efficacy outcome was VTE up to day 11, defined as deep vein thrombosis detected by mandatory bilateral venography or documented symptomatic deep vein thrombosis or pulmonary embolism. The primary safety outcome was major bleeding. RESULTS: Fondaparinux significantly reduced the incidence of VTE by day 11 (182 [6.8%] of 2682 patients) compared with enoxaparin (371 [13.7%] of 2703 patients), with a common odds reduction of 55.2% (95% confidence interval, 45.8% to 63.1%; $P < .001$); this beneficial effect was consistent across all types of surgery and all subgroups. Although major bleeding occurred more frequently in the fondaparinux-treated group ($P = .008$), the incidence of clinically relevant bleeding (leading to death or reoperation or occurring in a critical organ) did not differ between groups. CONCLUSION: In patients undergoing orthopedic surgery, 2.5 mg of fondaparinux sodium once daily, starting 6 hours postoperatively, showed a major benefit over enoxaparin, achieving an overall risk reduction of VTE greater than 50% without increasing the risk of clinically relevant bleeding.

Question: Therapy

Study Type: Meta-analysis

Learner Level: Beginner

Notes: Good to teach difference between a systematic review and a meta-analysis

Walters, J. A., et al. (2009). "Systemic corticosteroids for acute exacerbations of chronic obstructive pulmonary disease." Cochrane Database Syst Rev(1): CD001288.

BACKGROUND: COPD is a common condition, mainly related to smoking. Acute exacerbations of COPD, usually related to superimposed infection, occur commonly and systemic corticosteroids are widely used in their management in combination with other treatments including antibiotics, oxygen supplementation and bronchodilators. OBJECTIVES: To determine the efficacy of corticosteroids, administered either parenterally or orally, on the outcomes of acute exacerbations of COPD. SEARCH STRATEGY: Searches were carried out using the Cochrane Airways Group COPD RCT register with additional studies sought in the bibliographies of randomised controlled trials and review articles. Authors of identified randomised controlled trials were contacted for other published and unpublished studies. The last search was carried out in August 2008. SELECTION CRITERIA: Randomised controlled trials comparing corticosteroids, administered either parenterally or orally, with appropriate placebo control. Other interventions e.g. bronchodilators and antibiotics were standardised. Clinical studies of acute asthma were excluded. DATA COLLECTION AND ANALYSIS: Data were extracted independently by two reviewers. Data measured but not reported were sought from authors of included studies. Trials were combined using Review Manager for analyses. MAIN RESULTS: Eleven studies (n=1081) fulfilled the inclusion criteria and 10 studies contributed data for analyses (n=1051). There were significantly fewer treatment failures within thirty days in patients given corticosteroid treatment, Odds Ratio (OR) 0.50; 95% confidence interval (CI) 0.36 to 0.69 and Hazard Ratio 0.78; 95% CI 0.63 to 0.97. It would have been necessary to treat 10 patients (95%CI 7 to 16) with corticosteroids to avoid one treatment failure in this time period. Duration of hospitalisation was significantly shorter with corticosteroid treatment, mean difference -1.22 days; 95% CI -2.26 to -0.18. For FEV1 there were significant treatment benefits with mean differences at the early time point (to 72 hours), 140 ml; 95% CI 90 to 190 ml and at end of treatment (up to 15 days) 80 ml; 95% confidence interval 10 to 160. There was a significant improvement in breathlessness and blood gases at both time points. There was no significant effect on mortality but an increased likelihood of an adverse event associated with corticosteroid treatment, OR 2.33; 95% CI 1.60 to 3.40. Overall one extra adverse effect occurred for every 5 people treated (95% CI 4 to 9). The risk of hyperglycaemia was significantly increased, OR 4.95; 95% CI 2.47 to 9.91. AUTHORS' CONCLUSIONS: Treatment of an exacerbation of COPD with oral or parenteral corticosteroids significantly reduces treatment failure and the need for additional medical treatment and shortens hospital stay. It increases the rate of improvement in lung function and dyspnoea and the improvement continues during treatment, but there is a significantly increase in the risk of an adverse drug event occurring. The optimal dose and length of treatment regime needs to be better defined.

Question: Therapy

Study Type: Meta-analysis

Learner Level: Intermediate

Notes: Good article for review of the critical appraisal of a systematic review as methods are clearly reported. Teaching point can be made when discussing the outcome of 'treatment failure'. Page 8 describes the different meanings for this composite outcome in the various studies. The question to focus on: are the various outcomes in the composite measure similar in importance to the patient? (all cause mortality, readmission, intensification of pharmacologic treatment). Forest plots are clear and do have plots with both significant heterogeneity (figure 2 $I^2 = 43\%$) and others with $I^2 = 0$.

Wright, J. M. and V. M. Musini (2009). "First-line drugs for hypertension." *Cochrane Database Syst Rev*(3): CD001841.

BACKGROUND: Sustained elevated blood pressure, unresponsive to lifestyle measures, leads to a critically important clinical question: What class of drug to use first-line? This review answers that question. **OBJECTIVES:** Primary objective: To quantify the benefits and harms of the major first-line anti-hypertensive drug classes: thiazides, beta-blockers, calcium channel blockers, angiotensin converting enzyme (ACE) inhibitors, alpha-blockers, and angiotensin II receptor blockers (ARB). **SEARCH STRATEGY:** Electronic search of MEDLINE (Jan. 1966-June 2008), EMBASE, CINAHL, the Cochrane clinical trial register, using standard search strategy of the hypertension review group with additional terms. **SELECTION CRITERIA:** Randomized trials of at least one year duration comparing one of 6 major drug classes with a placebo or no treatment. More than 70% of people must have BP >140/90 mmHg at baseline. **DATA COLLECTION AND ANALYSIS:** The outcomes assessed were mortality, stroke, coronary heart disease (CHD), cardiovascular events (CVS), decrease in systolic and diastolic blood pressure, and withdrawals due to adverse drug effects. Risk ratio (RR) and a fixed effects model were used to combine outcomes across trials. **MAIN RESULTS:** Of 57 trials identified, 24 trials with 28 arms, including 58,040 patients met the inclusion criteria. Thiazides (19 RCTs) reduced mortality (RR 0.89, 95% CI 0.83, 0.96), stroke (RR 0.63, 95% CI 0.57, 0.71), CHD (RR 0.84, 95% CI 0.75, 0.95) and CVS (RR 0.70, 95% CI 0.66, 0.76). Low-dose thiazides (8 RCTs) reduced CHD (RR 0.72, 95% CI 0.61, 0.84), but high-dose thiazides (11 RCTs) did not (RR 1.01, 95% CI 0.85, 1.20). Beta-blockers (5 RCTs) reduced stroke (RR 0.83, 95% CI 0.72, 0.97) and CVS (RR 0.89, 95% CI 0.81, 0.98) but not CHD (RR 0.90, 95% CI 0.78, 1.03) or mortality (RR 0.96, 95% CI 0.86, 1.07). ACE inhibitors (3 RCTs) reduced mortality (RR 0.83, 95% CI 0.72-0.95), stroke (RR 0.65, 95% CI 0.52-0.82), CHD (RR 0.81, 95% CI 0.70-0.94) and CVS (RR 0.76, 95% CI 0.67-0.85). Calcium-channel blocker (1 RCT) reduced stroke (RR 0.58, 95% CI 0.41, 0.84) and CVS (RR 0.71, 95% CI 0.57, 0.87) but not CHD (RR 0.77 95% CI 0.55, 1.09) or mortality (RR 0.86 95% CI 0.68, 1.09). No RCTs were found for ARBs or alpha-blockers. **AUTHORS' CONCLUSIONS:** First-line low-dose thiazides reduce all morbidity and mortality outcomes. First-line ACE inhibitors and calcium channel blockers may be similarly effective but the evidence is less robust. First-line high-dose thiazides and first-line beta-blockers are inferior to first-line low-dose thiazides.

Question: Therapy

Study Type: Meta-analysis

Learner Level: Intermediate to Advanced

Notes: This is an extensive review of first line therapy for hypertension including studies comparing major drug classes with placebo/ no treatment looking at outcomes of mortality, stroke and coronary artery disease. Although this meta-analysis is very long (60 pages), a teaching session could focus on particular parts, such as specific forest plots. There is also an ACP journal club summary that could serve as a central point of discussion. The key messages here are that there are the gaps between what we know (thiazides and ACE inhibitors decrease mortality) and what we do (a minority of patients with hypertension are on these drugs as first line therapies).

Therapy

(1987). "Effects of enalapril on mortality in severe congestive heart failure. Results of the Cooperative North Scandinavian Enalapril Survival Study (CONSENSUS). The CONSENSUS Trial Study Group." *N Engl J Med* 316(23): 1429-1435.

To evaluate the influence of the angiotensin-converting-enzyme inhibitor enalapril (2.5 to 40 mg per day) on the prognosis of severe congestive heart failure (New York Heart Association [NYHA] functional class IV), we randomly assigned 253 patients in a double-blind study to receive either placebo (n = 126) or enalapril (n = 127). Conventional treatment for heart failure, including the use of other vasodilators, was continued in both groups. Follow-up averaged 188 days (range, 1 day to 20 months). The crude mortality at the end of six months (primary end point) was 26 percent in the enalapril group and 44 percent in the placebo group--a reduction of 40 percent (P = 0.002). Mortality was reduced by 31 percent at one year (P = 0.001). By the end of the study, there had been 68 deaths in the placebo group and 50 in the enalapril group--a reduction of 27 percent (P = 0.003). The entire reduction in total mortality was found to be among patients with progressive heart failure (a reduction of 50 percent), whereas no difference was seen in the incidence of sudden cardiac death. A significant improvement in NYHA classification was observed in the enalapril group, together with a reduction in heart size and a reduced requirement for other medication for heart failure. The overall withdrawal rate was similar in both groups, but hypotension requiring withdrawal occurred in seven patients in the enalapril group and in no patients in the placebo group. After the initial dose of enalapril was reduced to 2.5 mg daily in high-risk patients, this side effect was less frequent. We conclude that the addition of enalapril to conventional therapy in patients with severe congestive heart failure can reduce mortality and improve symptoms. The beneficial effect on mortality is due to a reduction in death from the progression of heart failure.

Question: Therapy

Study Type: RCT

Learner Level: Intermediate

Notes: This article has some historical significance in that it's the first demonstration of mortality benefit in heart failure with ACE inhibitors, and I like using it partly for that reason, as I think learners should have some idea about "landmark" papers. More significantly however, this is a great article for calculating number needed to treat (which ends up being very low) and then discussing why the number needed to treat is so low. In this study, the mortality was so high (because there was little other effective therapy and because the patients had such severe heart failure) that the intervention had a major impact, leading to such a low number needed to treat. Whenever I use this article, I usually then ask the group to tell me why we have to treat many more patients in current practice with ACE inhibitors to prevent one death, and it's because the underlying mortality of the condition has changed with other therapies now available.

(1997). "The effect of digoxin on mortality and morbidity in patients with heart failure. The Digitalis Investigation Group." *N Engl J Med* 336(8): 525-533.

BACKGROUND: The role of cardiac glycosides in treating patients with chronic heart failure and normal sinus rhythm remains controversial. We studied the effect of digoxin on mortality and hospitalization in a randomized, double-blind clinical trial. **METHODS:** In the main trial, patients with a left ventricular ejection fraction of 0.45 or less were randomly assigned to digoxin (3397 patients) or placebo (3403 patients) in addition to diuretics and angiotensin-converting-enzyme inhibitors (median dose of digoxin, 0.25 mg per day; average follow-up, 37 months). In an ancillary trial of patients with ejection fractions greater than 0.45, 492 patients were randomly assigned to digoxin and 496 to placebo. **RESULTS:** In the main trial, mortality was unaffected. There were 1181 deaths (34.8 percent) with digoxin and 1194 deaths (35.1 percent) with placebo (risk ratio when digoxin was compared with placebo, 0.99; 95 percent confidence interval, 0.91 to 1.07; P=0.80). In the digoxin group, there was a trend toward a decrease in the risk of death attributed to worsening heart failure (risk ratio, 0.88; 95 percent confidence interval, 0.77 to 1.01; P=0.06). There were 6 percent fewer hospitalizations overall in that group than in the placebo group, and fewer patients were hospitalized for worsening heart failure (26.8 percent vs. 34.7 percent; risk ratio, 0.72; 95 percent confidence interval, 0.66 to 0.79; P<0.001). In the ancillary trial, the findings regarding the primary combined outcome of death or hospitalization due to worsening heart failure were consistent with the results of the main trial. **CONCLUSIONS:** Digoxin did not reduce overall mortality, but it reduced the rate of hospitalization both overall and for worsening heart failure. These findings define more precisely the role of digoxin in the management of chronic heart failure.

Question: Therapy

Study Type: RCT

Learner Level: Beginner / Intermediate

Notes: This is a sample teaching package; Solid Methodology; Landmark paper in internal medicine; Important and controversial clinical question (Dig: to treat or not to treat?); Good discussion points: Can discuss both equivalency (outcomes that did not show difference) and also positive findings (can calculate an NNT- number needed to treat); Can discuss the balance of benefits of therapy vs. potential harms; Good clinical applicability discussion can follow; Does a 'good' paper ever grow old? (i.e. does publication date matter?).. if so, when?

(2002). "Major outcomes in high-risk hypertensive patients randomized to angiotensin-converting enzyme inhibitor or calcium channel blocker vs diuretic: The Antihypertensive and Lipid-Lowering Treatment to Prevent Heart Attack Trial (ALLHAT)." *JAMA* 288(23): 2981-2997.

CONTEXT: Antihypertensive therapy is well established to reduce hypertension-related morbidity and mortality, but the optimal first-step therapy is unknown. **OBJECTIVE:** To determine whether treatment with a calcium channel blocker or an angiotensin-converting enzyme inhibitor lowers the incidence of coronary heart disease (CHD) or other cardiovascular disease (CVD) events vs treatment with a diuretic. **DESIGN:** The Antihypertensive and Lipid-Lowering Treatment to Prevent Heart Attack Trial (ALLHAT), a randomized, double-blind, active-controlled clinical trial conducted from February 1994 through March 2002. **SETTING AND PARTICIPANTS:** A total of 33 357 participants aged 55 years or older with hypertension and at least 1 other CHD risk factor from 623 North American centers. **INTERVENTIONS:** Participants were randomly assigned to receive chlorthalidone, 12.5 to 25 mg/d (n = 15 255); amlodipine, 2.5 to 10 mg/d (n = 9048); or lisinopril, 10 to 40 mg/d (n = 9054) for planned follow-up of approximately 4 to 8 years. **MAIN OUTCOME MEASURES:** The primary outcome was combined fatal CHD or nonfatal myocardial infarction, analyzed by intent-to-treat. Secondary outcomes were all-cause mortality, stroke, combined CHD (primary outcome, coronary revascularization, or angina with hospitalization), and combined CVD (combined CHD, stroke, treated angina without hospitalization, heart failure [HF], and peripheral arterial disease). **RESULTS:** Mean follow-up was 4.9 years. The primary outcome occurred in 2956 participants, with no difference between treatments. Compared with chlorthalidone (6-year rate, 11.5%), the relative risks (RRs) were 0.98 (95% CI, 0.90-1.07) for amlodipine (6-year rate, 11.3%) and 0.99 (95% CI, 0.91-1.08) for lisinopril (6-year rate, 11.4%). Likewise, all-cause mortality did not differ between groups. Five-year systolic blood pressures were significantly higher in the amlodipine (0.8 mm Hg, P =.03) and lisinopril (2 mm Hg, P<.001) groups compared with chlorthalidone, and 5-year diastolic blood pressure was significantly lower with amlodipine (0.8 mm Hg, P<.001). For amlodipine vs chlorthalidone, secondary outcomes were similar except for a higher 6-year rate of HF with amlodipine (10.2% vs 7.7%; RR, 1.38; 95% CI, 1.25-1.52). For lisinopril vs chlorthalidone, lisinopril had higher 6-year rates of combined CVD (33.3% vs 30.9%; RR, 1.10; 95% CI, 1.05-1.16); stroke (6.3% vs 5.6%; RR, 1.15; 95% CI, 1.02-1.30); and HF (8.7% vs 7.7%; RR, 1.19; 95% CI, 1.07-1.31). **CONCLUSION:** Thiazide-type diuretics are superior in preventing 1 or more major forms of CVD and are less expensive. They should be preferred for first-step antihypertensive therapy.

Question: Therapy

Study Type: RCT

Learner Level: Very Advanced

Notes: This is an example of a very complicated paper with multiple arms, multiple outcomes and what feels like a zillion tables and graphs. However, it is also an example of a paper that had a great impact on internal medicine practice; Think of how you would approach this paper in a way that would not be intimidating, either as a practitioner who wants to understand the evidence in order to best apply it, or as an educator who would like to summarize the most important points so the importance of this study can be appreciated. Warning: This paper is not for the faint at heart and the program is not responsible for any unintended harms that may come from reading this paper.

(2011). "Acetylcysteine for prevention of renal outcomes in patients undergoing coronary and peripheral vascular angiography: main results from the randomized Acetylcysteine for Contrast-induced nephropathy Trial (ACT)." *Circulation* 124(11): 1250-1259.

BACKGROUND: It remains uncertain whether acetylcysteine prevents contrast-induced acute kidney injury. **METHODS AND RESULTS:** We randomly assigned 2308 patients undergoing an intravascular angiographic procedure with at least 1 risk factor for contrast-induced acute kidney injury (age >70 years, renal failure, diabetes mellitus, heart failure, or hypotension) to acetylcysteine 1200 mg or placebo. The study drugs were administered orally twice daily for 2 doses before and 2 doses after the procedure. The allocation was concealed (central Web-based randomization). All analysis followed the intention-to-treat principle. The incidence of contrast-induced acute kidney injury (primary end point) was 12.7% in the acetylcysteine group and 12.7% in the control group (relative risk, 1.00; 95% confidence interval, 0.81 to 1.25; P=0.97). A combined end point of mortality or need for dialysis at 30 days was also similar in both groups (2.2% and 2.3%, respectively; hazard ratio, 0.97; 95% confidence interval, 0.56 to 1.69; P=0.92). Consistent effects were observed in all subgroups analyzed, including those with renal impairment. **CONCLUSIONS:** In this large randomized trial, we found that acetylcysteine does not reduce the risk of contrast-induced acute kidney injury or other clinically relevant outcomes in at-risk patients undergoing coronary and peripheral vascular angiography. **CLINICAL TRIAL REGISTRATION:** <http://www.clinicaltrials.gov>. Unique identifier: NCT00736866.

Question: Therapy

Study Type: RCT

Learner Level: Beginner

Notes: Beginner level therapy paper.

Abell, T., et al. (2003). "Gastric electrical stimulation for medically refractory gastroparesis." *Gastroenterology* 125(2): 421-428.

BACKGROUND & AIMS: This study investigated the efficacy of gastric electrical stimulation for the treatment of symptomatic gastroparesis unresponsive to standard medical therapy. **METHODS:** Thirty-three patients with chronic gastroparesis (17 diabetic and 16 idiopathic) received continuous high-frequency/low-energy gastric electrical stimulation via electrodes in the muscle wall of the antrum connected to a neurostimulator in an abdominal wall pocket. After implantation, patients were randomized in a double-blind crossover design to stimulation ON or OFF for 1-month periods. The blind was then broken, and all patients were programmed to stimulation ON and evaluated at 6 and 12 months. Outcome measures were vomiting frequency, preference for ON or OFF, upper gastrointestinal tract symptoms, quality of life, gastric emptying, and adverse events. **RESULTS:** In the double-blind portion of the study, self-reported vomiting frequency was significantly reduced in the ON vs. OFF period (P < 0.05) and this symptomatic improvement was consistent with the significant patient preference (P < 0.05) for the ON vs. OFF period determined before breaking the blind. In the unblinded portion of the study, vomiting frequency decreased significantly (P < 0.05) at 6 and 12 months. Scores for symptom severity and quality of life significantly improved (P < 0.05) at 6 and 12 months, whereas gastric emptying was only modestly accelerated. Five patients had their gastric electrical stimulation system explanted or revised because of infection or other complications. **CONCLUSIONS:** High-frequency/low-energy gastric electrical stimulation significantly decreased vomiting frequency and gastrointestinal symptoms and improved quality of life in patients with severe gastroparesis.

Question: Therapy

Study Type: RCT

Learner Level: Beginner

Notes: Good to teach NNT/ARR

Amarencu, P., et al. (2006). "High-dose atorvastatin after stroke or transient ischemic attack." *N Engl J Med* 355(6): 549-559.

BACKGROUND: Statins reduce the incidence of strokes among patients at increased risk for cardiovascular disease; whether they reduce the risk of stroke after a recent stroke or transient ischemic attack (TIA) remains to be established. **METHODS:** We randomly assigned 4731 patients who had had a stroke or TIA within one to six months before study entry, had low-density lipoprotein (LDL) cholesterol levels of 100 to 190 mg per deciliter (2.6 to 4.9 mmol per liter), and had no known coronary heart disease to double-blind treatment with 80 mg of atorvastatin per day or placebo. The primary end point was a first nonfatal or fatal stroke. **RESULTS:** The mean LDL cholesterol level during the trial was 73 mg per deciliter (1.9 mmol per liter) among patients receiving atorvastatin and 129 mg per deciliter (3.3 mmol per liter) among patients receiving placebo. During a median follow-up of 4.9 years, 265 patients

(11.2 percent) receiving atorvastatin and 311 patients (13.1 percent) receiving placebo had a fatal or nonfatal stroke (5-year absolute reduction in risk, 2.2 percent; adjusted hazard ratio, 0.84; 95 percent confidence interval, 0.71 to 0.99; $P=0.03$; unadjusted $P=0.05$). The atorvastatin group had 218 ischemic strokes and 55 hemorrhagic strokes, whereas the placebo group had 274 ischemic strokes and 33 hemorrhagic strokes. The five-year absolute reduction in the risk of major cardiovascular events was 3.5 percent (hazard ratio, 0.80; 95 percent confidence interval, 0.69 to 0.92; $P=0.002$). The overall mortality rate was similar, with 216 deaths in the atorvastatin group and 211 deaths in the placebo group ($P=0.98$), as were the rates of serious adverse events. Elevated liver enzyme values were more common in patients taking atorvastatin. **CONCLUSIONS:** In patients with recent stroke or TIA and without known coronary heart disease, 80 mg of atorvastatin per day reduced the overall incidence of strokes and of cardiovascular events, despite a small increase in the incidence of hemorrhagic stroke. (ClinicalTrials.gov number, NCT00147602 [ClinicalTrials.gov].)

Question: Therapy

Study Type: RCT

Learner Level: Beginner

Notes: Good article that will allow for conversation regarding stroke and secondary prevention. Use Table 2 in order to demonstrate how to draw a 2x2 table of the outcomes and discuss calculating ARR and NNT.

Annane, D., et al. (2004). "Corticosteroids for severe sepsis and septic shock: a systematic review and meta-analysis." *BMJ* 329(7464): 480.

OBJECTIVE: To assess the effects of corticosteroids on mortality in patients with severe sepsis and septic shock. **DATA SOURCES:** Randomised and quasi-randomised trials of corticosteroids versus placebo (or supportive treatment alone) retrieved from the Cochrane infectious diseases group's trials register, the Cochrane central register of controlled trials, Medline, Embase, and LILACS. **REVIEW METHOD:** Two pairs of reviewers agreed on eligibility of trials. One reviewer entered data on to the computer and four reviewers checked them. We obtained some missing data from authors of trials and assessed methodological quality of trials. **RESULTS:** 16/23 trials ($n = 2063$) were selected. Corticosteroids did not change 28 day mortality (15 trials, $n = 2022$; relative risk 0.92, 95% confidence interval 0.75 to 1.14) or hospital mortality (13 trials, $n = 1418$; 0.89, 0.71 to 1.11). There was significant heterogeneity. Subgroup analysis on long courses ($> \text{ or } = 5$ days) with low dose ($< \text{ or } = 300$ mg hydrocortisone or equivalent) corticosteroids showed no more heterogeneity. The relative risk for mortality was 0.80 at 28 days (five trials, $n = 465$; 0.67 to 0.95) and 0.83 at hospital discharge (five trials, $n = 465$, 0.71 to 0.97). Use of corticosteroids reduced mortality in intensive care units (four trials, $n = 425$, 0.83, 0.70 to 0.97), increased shock reversal at 7 days (four trials, $n = 425$; 1.60, 1.27 to 2.03) and 28 days (four trials, $n = 425$, 1.26, 1.04 to 1.52) without inducing side effects. **CONCLUSIONS:** For all trials, regardless of duration of treatment and dose, use of corticosteroids did not significantly affect mortality. With long courses of low doses of corticosteroids, however, mortality at 28 days and hospital mortality was reduced.

Question: Therapy

Study Type: Meta-analysis

Learner Level: Intermediate / Advanced

Notes: This is a well-done meta-analysis that can allow discussion of important concepts for intermediate to advanced learners who want to push their understanding of meta-analysis. ACP Journal Club Summary 2005; 142 (2): 30. Good discussion points: Can discuss both quasi-randomization (definition: pseudo random allocation method such as day of birth, order of participant enrollment in study, day of the week, medical record number etc.) and the introduction of possible selection bias. Can refer readers to tables 3 and 4 where specific design issues are identified. Can also discuss heterogeneity and the authors' attempts to explain it as multifactorial, including differences in methodologic strength and dose/ duration of intervention. Excellent review of how to think about subgroup analysis. Possible strengths (pre-stated hypothesis, statistically significant effect), and pitfalls (difference is comparisons between studies rather than within studies) of the interpretation of this subgroup analysis.

April, M. D., et al. (2018). "Aromatherapy Versus Oral Ondansetron for Antiemetic Therapy Among Adult Emergency Department Patients: A Randomized Controlled Trial." *Ann Emerg Med*.

STUDY OBJECTIVE: We compare aromatherapy with inhaled isopropyl alcohol versus oral ondansetron for treating nausea among emergency department (ED) patients not requiring immediate intravenous access. **METHODS:** In a randomized, blinded, placebo-controlled trial, we enrolled a convenience sample of adults presenting to an urban tertiary care ED with chief complaints including nausea or vomiting. We randomized subjects to 1 of 3 arms: inhaled isopropyl alcohol and 4 mg oral ondansetron, inhaled isopropyl alcohol and oral placebo, and inhaled saline solution placebo and 4 mg oral ondansetron. The primary outcome was mean nausea reduction measured by a 0- to 100-mm visual analog scale from enrollment to 30 minutes postintervention. Secondary outcomes included receipt of rescue antiemetic medications and adverse events. **RESULTS:** We enrolled 122 subjects, of whom 120 (98.3%) completed the study. Of randomized subjects, 40 received inhaled isopropyl alcohol and oral ondansetron, 41 received inhaled isopropyl alcohol and oral placebo, and 41 received inhaled saline solution placebo and oral ondansetron. The mean decrease in nausea visual analog scale score in each arm was 30 mm (95% confidence interval [CI] 22 to 37 mm), 32 mm (95% CI 25 to 39 mm), and 9 mm (95% CI 5 to 14 mm), respectively. The proportions of subjects who received rescue antiemetic therapy in each arm were 27.5% (95% CI 14.6% to 43.9%), 25.0% (95% CI 12.7% to 41.2%), and 45.0% (95% CI 29.3% to 61.5%), respectively. There were no adverse events. **CONCLUSION:** Among ED patients with acute nausea and not requiring immediate intravenous access, aromatherapy with or without oral ondansetron provides greater nausea relief than oral ondansetron alone.

Question: Therapy

Study Type: RCT

Learner Level: beginner/intermediate

Notes: Very straight forward RCT, which is sometimes difficult to find

Austin, M. A., et al. (2010). "Effect of high flow oxygen on mortality in chronic obstructive pulmonary disease patients in prehospital setting: randomised controlled trial." *BMJ* 341: c5462.

OBJECTIVES: To compare standard high flow oxygen treatment with titrated oxygen treatment for patients with an acute exacerbation of chronic obstructive pulmonary disease in the prehospital setting. **DESIGN:** Cluster randomised controlled parallel group trial. **SETTING:** Ambulance service in Hobart, Tasmania, Australia. **PARTICIPANTS:** 405 patients with a presumed acute exacerbation of chronic obstructive pulmonary disease who were treated by paramedics, transported, and admitted to the Royal Hobart Hospital during the trial period; 214 had a diagnosis of chronic obstructive pulmonary disease confirmed by lung function tests in the previous five years. **INTERVENTIONS:** High flow oxygen treatment compared with titrated oxygen treatment in the prehospital (ambulance/paramedic) setting. **MAIN OUTCOME MEASURE:** Prehospital or in-hospital mortality. **RESULTS:** In an intention to treat analysis, the risk of death was significantly lower in the titrated oxygen arm compared with the high flow oxygen arm for all patients (high flow oxygen n=226; titrated oxygen n=179) and for the subgroup of patients with confirmed chronic obstructive pulmonary disease (high flow n=117; titrated n=97). Overall mortality was 9% (21 deaths) in the high flow oxygen arm compared with 4% (7 deaths) in the titrated oxygen arm; mortality in the subgroup with confirmed chronic obstructive pulmonary disease was 9% (11 deaths) in the high flow arm compared with 2% (2 deaths) in the titrated oxygen arm. Titrated oxygen treatment reduced mortality compared with high flow oxygen by 58% for all patients (relative risk 0.42, 95% confidence interval 0.20 to 0.89; P=0.02) and by 78% for the patients with confirmed chronic obstructive pulmonary disease (0.22, 0.05 to 0.91; P=0.04). Patients with chronic obstructive pulmonary disease who received titrated oxygen according to the protocol were significantly less likely to have respiratory acidosis (mean difference in pH 0.12 (SE 0.05); P=0.01; n=28) or hypercapnia (mean difference in arterial carbon dioxide pressure -33.6 (16.3) mm Hg; P=0.02; n=29) than were patients who received high flow oxygen. **CONCLUSIONS:** Titrated oxygen treatment significantly reduced mortality, hypercapnia, and respiratory acidosis compared with high flow oxygen in acute exacerbations of chronic obstructive pulmonary disease. These results provide strong evidence to recommend the routine use of titrated oxygen treatment in patients with breathlessness and a history or clinical likelihood of chronic obstructive pulmonary disease in the prehospital setting. **TRIAL REGISTRATION:** Australian New Zealand Clinical Trials Register ACTRN12609000236291.

Question: Therapy

Study Type: RCT

Learner Level: .

Notes: COPD and titrated O2: More and more studies of complex interventions will use a cluster randomized design. Savvy evidence consumers need to know the advantages and limitations of these kinds of studies and how to appraise them.

Bajwa, E. K., et al. (2005). "Interferon-gamma1b therapy in idiopathic pulmonary fibrosis: a metaanalysis." *Chest* 128(1): 203-206.

CONTEXT: Despite the investigation of multiple therapeutic options, idiopathic pulmonary fibrosis (IPF) remains a devastating, progressively fatal disease. Much interest has focused on the use of interferon (IFN)-gamma1b therapy, but the efficacy of this treatment has not been proven. **OBJECTIVE:** To determine whether IFN treatment reduces mortality in patients with IPF. **DESIGN:** A meta-analysis of randomized controlled trials evaluating the use of IFN-gamma1b as treatment for IPF. **MAIN OUTCOME MEASURE:** Mortality in patients treated with IFN-gamma1b was compared to mortality in patients treated with control therapies. **RESULTS:** A total of three studies involving 390 patients was included in the analysis. IFN-gamma1b therapy was associated with reduced mortality (hazard ratio [HR], 0.418; 95% confidence interval [CI], 0.253 to 0.690; p = 0.0003). A comparison of mortality at different time points revealed that IFN-gamma1b therapy was associated with significantly reduced mortality at 1 year (0.0861; 95% CI, 0.0244 to 0.1478; p = 0.0063), 18 months (0.1682; 95% CI, 0.1065 to 0.2299; p < 0.0001), 650 days (0.1939; 95% CI, 0.1386 to 0.2492; p < 0.0001), and 2 years (0.2652; 95% CI, 0.1652 to 0.3652; p < 0.0001). **CONCLUSION:** When the results of multiple studies are combined in a meta-analysis, IFN-gamma1b therapy is associated with reduced mortality.

Question: Therapy

Study Type: Meta-analysis

Learner Level: Intermediate

Notes: This can be taught as a package. First exercise goes through the user's guide for therapy and the first article. The second exercise goes through the user's guide for meta-analysis. It helps illustrate some of the points of the validity of the systematic review because the learner's have had a sneak peak at the underlying data. Teach with: Part 1: Raghu G et al. A Placebo-controlled trial of interferon gamma 1b in patients with idiopathic pulmonary fibrosis

Baker, W. L., et al. (2009). "Systematic review: comparative effectiveness of angiotensin-converting enzyme inhibitors or angiotensin II-receptor blockers for ischemic heart disease." *Ann Intern Med* 151(12): 861-871.

BACKGROUND: Patients with ischemic heart disease and preserved ventricular function experience considerable morbidity and mortality despite standard medical therapy. **PURPOSE:** To compare benefits and harms of using angiotensin-converting enzyme (ACE) inhibitors, angiotensin II-receptor blockers (ARBs), or combination therapy in adults with stable ischemic heart disease and preserved ventricular function. **DATA SOURCES:** MEDLINE, EMBASE, Cochrane Central Register of Controlled Trials, and Cochrane Database of Systematic Reviews (earliest date, July 2009) were searched without language restrictions. **STUDY SELECTION:** Two independent investigators screened citations for trials of at least 6 months' duration that compared ACE inhibitors, ARBs, or combination therapy with placebo or active control and reported any of several clinical outcomes. **DATA EXTRACTION:** Using standardized protocols, 2 independent investigators extracted information about study characteristics and rated the quality and strength of evidence. Disagreement was resolved by consensus. **DATA SYNTHESIS:** 41 studies met eligibility criteria. Moderate- to high-strength evidence (7 trials; 32 559 participants) showed that ACE inhibitors reduce the relative risk (RR) for total mortality (RR, 0.87 [95% CI, 0.81 to 0.94]) and nonfatal myocardial infarction (RR, 0.83 [CI, 0.73 to 0.94]) but increase the RR for syncope (RR, 1.24 [CI, 1.02 to 1.52]) and cough (RR, 1.67 [CI, 1.22 to 2.29]) compared with placebo. Low-strength evidence (1 trial; 5926 participants) suggested that ARBs reduce the RR for the composite end point of cardiovascular mortality, nonfatal myocardial infarction, or stroke (RR, 0.88 [CI, 0.77 to 1.00]) but not for the individual components. Moderate-strength evidence (1 trial; 25 620 participants) showed similar effects on total mortality (RR, 1.07 [CI, 0.98 to 1.16]) and myocardial infarction (RR, 1.08 [CI, 0.94 to 1.23]) but an increased risk for discontinuations because of hypotension ($P < 0.001$) and syncope ($P = 0.035$) with combination therapy compared with ACE inhibitors alone. **LIMITATIONS:** Many studies either did not assess or did not report harms in a systematic manner. Many studies did not adequately report benefits or harms by various patient subgroups. **CONCLUSION:** Adding an ACE inhibitor to standard medical therapy improves outcomes, including reduced risk for mortality and myocardial infarctions, in some patients with stable ischemic heart disease and preserved ventricular function. Less evidence supports a benefit of ARB therapy, and combination therapy seems no better than ACE inhibitor therapy alone and increases harms. **PRIMARY FUNDING SOURCE:** Agency for Healthcare Research and Quality.

Question: Therapy

Study Type: Systematic Review

Learner Level: Beginner

Notes: Well done meta-analysis; Nicely reported with transparent grading of evidence; Benefits nicely documented, but harms less so; Forest plots can be used to teach heterogeneity (Figure A. total mortality)

Bangalore, S., et al. (2008). "Perioperative beta blockers in patients having non-cardiac surgery: a meta-analysis." *Lancet* 372(9654): 1962-1976.

BACKGROUND: American College of Cardiology and American Heart Association (ACC/AHA) guidelines on perioperative assessment recommend perioperative beta blockers for non-cardiac surgery, although results of some clinical trials seem not to support this recommendation. We aimed to critically review the evidence to assess the use of perioperative beta blockers in patients having non-cardiac surgery. **METHODS:** We searched Pubmed and Embase for randomised controlled trials investigating the use of beta blockers in non-cardiac surgery. We extracted data for 30-day all-cause mortality, cardiovascular mortality, non-fatal myocardial infarction, non-fatal stroke, heart failure, and myocardial ischaemia, safety outcomes of perioperative bradycardia, hypotension, and bronchospasm. **FINDINGS:** 33 trials included 12 306 patients. beta blockers were not associated with any significant reduction in the risk of all-cause mortality, cardiovascular mortality, or heart failure, but were associated with a decrease (odds ratio [OR] 0.65, 95% CI 0.54-0.79) in non-fatal myocardial infarction (number needed to treat [NNT] 63) and decrease (OR 0.36, 0.26-0.50) in myocardial ischaemia (NNT 16) at the expense of an increase (OR 2.01, 1.27-3.68) in non-fatal strokes (number needed to harm [NNH] 293). The beneficial effects were driven mainly by trials with high risk of bias. For the safety outcomes, beta blockers were associated with a high risk of perioperative bradycardia requiring treatment (NNH 22), and perioperative hypotension requiring treatment (NNH 17). We recorded no increased risk of bronchospasm. **INTERPRETATION:** Evidence does not support the use of beta-blocker therapy for the prevention of perioperative clinical outcomes in patients having non-cardiac surgery. The ACC/AHA guidelines committee should soften their advocacy for this intervention until conclusive evidence is available.

Question: Therapy

Study Type: Meta-Analysis

Learner Level: .

Notes: MA

Becker, G., et al. (2006). "Malignant ascites: systematic review and guideline for treatment." *Eur J Cancer* 42(5): 589-597.

A guideline on the management of symptomatic malignant ascites by abdominal paracentesis, diuretics and peritoneovenous shunting, based on a systematic review of the literature is presented. Thirty-two relevant studies were identified. None were randomized control trials, one was a non-randomized open controlled trial, five were cohort studies or prospective uncontrolled trials, 26 studies were non-analytic studies like case series. Although paracentesis, diuretics and shunting are commonly used procedures, the evidence is weak. Available data show good, although temporary effect of paracentesis on symptom relief. Fluid withdrawal speed and concurrent intravenous hydration is not sufficiently studied. Peritoneovenous shunts can control ascites in patients with malignant ascites, but have to be balanced by the potential risks of this procedure. The available data about diuretics in treatment of malignant ascites are controversial. The use of diuretics therefore should be considered in all patients, but has to be evaluated individually.

Question: Therapy

Study Type: Systematic Review

Learner Level: Beginner

Notes: Good to teach difference between a meta-analysis and a systematic review

Bent, S., et al. (1999). "Antibiotics in acute bronchitis: a meta-analysis." *Am J Med* 107(1): 62-67.

PURPOSE: Most patients with acute bronchitis who seek medical care are treated with antibiotics, although the effectiveness of this intervention is uncertain. We performed a meta-analysis of randomized, controlled trials to estimate the effectiveness of antibiotics in the treatment of acute bronchitis. **SUBJECTS AND METHODS:** English-language studies published January 1966 to April 1998 were retrieved using MEDLINE, bibliographies, and consultation with experts. Only randomized trials that enrolled otherwise healthy patients with a diagnosis of acute bronchitis, used an antibiotic in the treatment group and a placebo in the control group, and provided sufficient data to calculate an effect size were included. **RESULTS:** We identified eight randomized controlled trials that satisfied all inclusion criteria. These studies used one of three antibiotics (erythromycin, doxycycline, trimethoprim/sulfamethoxazole). The use of antibiotics decreased the duration of cough and sputum production by approximately one-half day (summary effect size 0.21; 95% CI, 0.05 to 0.36). For specific symptoms, there were nonsignificant trends favoring the use of antibiotics: a decrease of 0.4 days of purulent sputum (95% CI, -0.1 to 0.8), a decrease of 0.5 days of cough (95% CI, -0.1 to 1.1), and a decrease of 0.3 days lost from work (95% CI, -0.6 to 1.1). **CONCLUSION:** This meta-analysis suggests a small benefit from the use of the antibiotics erythromycin, doxycycline, or trimethoprim/sulfamethoxazole in the treatment of acute bronchitis in otherwise healthy patients. As this small benefit must be weighed against the risk of side effects and the societal cost of increasing antibiotic resistance, we believe that the use of antibiotics is not justified in these patients.

Question: Therapy

Study Type: Systematic Review

Learner Level: Intermediate

Notes: There are nice tables and figures that are classic for systematic review and meta-analysis. Good discussion points: This is a good paper for discussion of 'effect size,' (what you do when the papers in your study don't measure the same outcomes..) However, this may slightly confuse those who are not fully comfortable with meta-analysis.

Bischoff-Ferrari, H. A., et al. (2009). "Fall prevention with supplemental and active forms of vitamin D: a meta-analysis of randomised controlled trials." *BMJ* 339: b3692.

OBJECTIVE: To test the efficacy of supplemental vitamin D and active forms of vitamin D with or without calcium in preventing falls among older individuals. **DATA SOURCES:** We searched Medline, the Cochrane central register of controlled trials, BIOSIS, and Embase up to August 2008 for relevant articles. Further studies were identified by consulting clinical experts, bibliographies, and abstracts. We contacted authors for additional data when necessary. **Review methods** Only double blind randomised controlled trials of older individuals (mean age 65 years or older) receiving a defined oral dose of supplemental vitamin D (vitamin D(3) (cholecalciferol) or vitamin D(2) (ergocalciferol)) or an active form of vitamin D (1alpha-hydroxyvitamin D(3) (1alpha-hydroxycalciferol) or 1,25-dihydroxyvitamin D(3) (1,25-dihydroxycholecalciferol)) and with sufficiently specified fall assessment were considered for inclusion. **RESULTS:** Eight randomised controlled trials (n=2426) of supplemental vitamin D met our inclusion criteria. Heterogeneity among trials was observed for dose of vitamin D (700-1000 IU/day v 200-600 IU/day; P=0.02) and achieved 25-hydroxyvitamin D(3) concentration (25(OH)D concentration: <60 nmol/l v >or=60 nmol/l; P=0.005). High dose supplemental vitamin D reduced fall risk by 19% (pooled relative risk (RR) 0.81, 95% CI 0.71 to 0.92; n=1921 from seven trials), whereas achieved serum 25(OH)D concentrations of 60 nmol/l or more resulted in a 23% fall reduction (pooled RR 0.77, 95% CI 0.65 to 0.90). Falls were not notably reduced by low dose supplemental vitamin D (pooled RR 1.10, 95% CI 0.89 to 1.35; n=505 from two trials) or by achieved serum 25-hydroxyvitamin D concentrations of less than 60 nmol/l (pooled RR 1.35, 95% CI 0.98 to 1.84). Two randomised controlled trials (n=624) of active forms of vitamin D met our inclusion criteria. Active forms of vitamin D reduced fall risk by 22% (pooled RR 0.78, 95% CI 0.64 to 0.94). **CONCLUSIONS:** Supplemental vitamin D in a dose of 700-1000 IU a day reduced the risk of falling among older individuals by 19% and to a similar degree as active forms of vitamin D. Doses of supplemental vitamin D of less than 700 IU or serum 25-hydroxyvitamin D concentrations of less than 60 nmol/l may not reduce the risk of falling among older individuals.

Question: Therapy

Study Type: Meta-analysis

Learner Level: Intermediate

Notes: Nicely reported meta-analysis that can be used to illustrate issues of heterogeneity. When all studies were included in the combined outcome, the results showed significant heterogeneity. However, if the data are separated into high dose and low dose, the heterogeneity is significantly improved. There is an ACP Journal club summary for this article.

Boden, W. E., et al. (2011). "Niacin in patients with low HDL cholesterol levels receiving intensive statin therapy." *N Engl J Med* 365(24): 2255-2267.

BACKGROUND: In patients with established cardiovascular disease, residual cardiovascular risk persists despite the achievement of target low-density lipoprotein (LDL) cholesterol levels with statin therapy. It is unclear whether extended-release niacin added to simvastatin to raise low levels of high-density lipoprotein (HDL) cholesterol is superior to simvastatin alone in reducing such residual risk. **METHODS:** We randomly assigned eligible patients to receive extended-release niacin, 1500 to 2000 mg per day, or matching placebo. All patients received simvastatin, 40 to 80 mg per day, plus ezetimibe, 10 mg per day, if needed, to maintain an LDL cholesterol level of 40 to 80 mg per deciliter (1.03 to 2.07 mmol per liter). The primary end point was the first event of the composite of death from coronary heart disease, nonfatal myocardial infarction, ischemic stroke, hospitalization for an acute coronary syndrome, or symptom-driven coronary or cerebral revascularization. **RESULTS:** A total of 3414 patients were randomly assigned to receive niacin (1718) or placebo (1696). The trial was stopped after a mean follow-up period of 3 years owing to a lack of efficacy. At 2 years, niacin therapy had significantly increased the median HDL cholesterol level from 35 mg per deciliter (0.91 mmol per liter) to 42 mg per deciliter (1.08 mmol per liter), lowered the triglyceride level from 164 mg per deciliter (1.85 mmol per liter) to 122 mg per deciliter (1.38 mmol per liter), and lowered the LDL cholesterol level from 74 mg per deciliter (1.91 mmol per liter) to 62 mg per deciliter (1.60 mmol per liter). The primary end point occurred in 282 patients in the niacin group (16.4%) and in 274 patients in the placebo group (16.2%) (hazard ratio, 1.02; 95% confidence interval, 0.87 to 1.21; P=0.79 by the log-rank test). **CONCLUSIONS:** Among patients with atherosclerotic cardiovascular disease and LDL cholesterol levels of less than 70 mg per deciliter (1.81 mmol per liter), there was no incremental clinical benefit from the addition of niacin to statin therapy during a 36-month follow-up period, despite significant improvements in HDL cholesterol and triglyceride levels. (Funded by the National Heart, Lung, and Blood Institute and Abbott Laboratories; AIM-HIGH ClinicalTrials.gov number, NCT00120289.)

Question: Therapy

Study Type: RCT

Learner Level: Beginner

Notes: AIM-HIGH trial, negative niacin trial - great for teaching about composite outcomes; trials stopped early and loss of power; surrogate outcomes (HDL vs. clinical outcomes).

Bradley, E. H., et al. (2001). "A qualitative study of increasing beta-blocker use after myocardial infarction: Why do some hospitals succeed?" JAMA 285(20): 2604-2611.

CONTEXT: Based on evidence that beta-blockers can reduce mortality in patients with acute myocardial infarction (AMI), many hospitals have initiated performance improvement efforts to increase prescription of beta-blockers at discharge. Determination of the factors associated with such improvements may provide guidance to hospitals that have been less successful in increasing beta-blocker use. **OBJECTIVES:** To identify factors that may influence the success of improvement efforts to increase beta-blocker use after AMI and to develop a taxonomy for classifying such efforts. **DESIGN, SETTING, AND PARTICIPANTS:** Qualitative study in which data were gathered from in-depth interviews conducted in March-June 2000 with 45 key physician, nursing, quality management, and administrative participants at 8 US hospitals chosen to represent a range of hospital sizes, geographic regions, and changes in beta-blocker use rates between October 1996 and September 1999. **MAIN OUTCOME MEASURES:** Initiatives, strategies, and approaches to improve care for patients with AMI. **RESULTS:** The interviews revealed 6 broad factors that characterized hospital-based improvement efforts: goals of the efforts, administrative support, support among clinicians, design and implementation of improvement initiatives, use of data, and modifying variables. Hospitals with greater improvements in beta-blocker use over time demonstrated 4 characteristics not found in hospitals with less or no improvement: shared goals for improvement, substantial administrative support, strong physician leadership advocating beta-blocker use, and use of credible data feedback. **CONCLUSIONS:** This study provides a context for understanding efforts to improve care in the hospital setting by describing a taxonomy for classifying and evaluating such efforts. In addition, the study suggests possible elements of successful efforts to increase beta-blocker use for patients with AMI.

Question: Therapy

Study Type: Qualitative / In-depth interviews

Learner Level: Beginner / Intermediate

Notes: Excellent study for consideration of how qualitative research is done and what it can contribute to the medical literature that is distinct from quantitative research

Brandjes, D. P., et al. (1997). "Randomised trial of effect of compression stockings in patients with symptomatic proximal-vein thrombosis." Lancet 349(9054): 759-762.

BACKGROUND: Post-thrombotic syndrome varies from mild oedema to incapacitating swelling with pain and ulceration. We investigated the rate of post-thrombotic syndrome after a first episode of deep-vein thrombosis and assessed the preventive effect of direct application of a sized-to-fit graded compression stocking. **METHODS:** Patients with a first episode of venogram-proven proximal deep-vein thrombosis were randomly assigned no stockings (the control group) or made-to-measure graded compression elastic stockings for at least 2 years. Post-thrombotic syndrome was assessed with a standard scoring system that combined clinical characteristics and objective leg measurements. Patients were assessed every 3 months during the first 2 years, and every 6 months thereafter for at least 5 years. The cumulative incidence of mild-to-moderate post-thrombotic syndrome was the primary outcome measure. **FINDINGS:** Of the 315 consecutive outpatients considered for inclusion, 44 were excluded and 77 did not consent to take part. 194 patients were randomly assigned compression stockings (n = 96) or no stockings (n = 98). The median follow-up was 76

months (range 60-96) in both groups. Mild-to-moderate post-thrombotic syndrome (score ≥ 3 plus one clinical sign) occurred in 19 (20%) patients in the stocking group and in 46 (47%) control-group patients ($p < 0.001$). 11 (11%) patients in the stocking group developed severe post-thrombotic syndrome (score ≥ 4), compared with 23 (23%) patients in the control group ($p < 0.001$). In both groups, most cases of post-thrombotic syndrome occurred within 24 months of the acute thrombotic event. INTERPRETATION: About 60% of patients with a first episode of proximal deep-vein thrombosis develop post-thrombotic syndrome within 2 years. A sized-to-fit compression stocking reduced this rate by about 50%.

Question: Therapy

Study Type: RCT

Learner Level: Beginner / Intermediate

Notes: Good, straightforward RCT methods; Easy NNT (number needed to treat) calculations. Good discussion points: Discussion of the strengths and weaknesses of subjective and objective outcomes; Can discuss the difference between time sensitive analysis (figure 2) as compared with proportions of patients who have responded at a particular time. Those wishing to take on a slightly greater challenge can discuss Kaplan-Meier Analysis (page 760 and figure 2)

Bruckert, E., et al. (2010). "Meta-analysis of the effect of nicotinic acid alone or in combination on cardiovascular events and atherosclerosis." *Atherosclerosis* 210(2): 353-361.

OBJECTIVE: High-density lipoprotein cholesterol (HDL-C) concentration is a strong predictor of cardiovascular events in both naive and statin-treated patients. Nicotinic acid is an attractive option for decreasing residual risk in statin-treated or statin-intolerant patients since it increases HDL-C by up to 20% and decreases low-density lipoprotein cholesterol and lipoprotein(a) plasma concentrations. **METHODS:** We performed a computerized PubMed literature search that focused on clinical trials evaluating niacin, alone or in combination with other lipid-lowering drugs, published between January 1966 and August 2008. **RESULTS:** Among 587 citations, 29 full articles were read and 14 were eligible for inclusion. Overall 11 randomized controlled trials enrolled 2682 patients in the active group and 3934 in the control group. In primary analysis, niacin significantly reduced major coronary events (relative odds reduction=25%, 95% CI 13, 35), stroke (26%, 95% CI=8, 41) and any cardiovascular events (27%, 95% CI=15, 37). Except for stroke, the pooled between-group difference remained significant in sensitivity analysis excluding the largest trial. In comparison with the non-niacin group, more patients in the niacin group had regression of coronary atherosclerosis (relative increase=92%, 95% CI=39, 67) whereas the rate of patients with progression decreased by 41%, 95% CI=25, 53. Similar effects of niacin were found on carotid intima thickness with a weighted mean difference in annual change of -17 microm/year (95% CI=-22, -12). **CONCLUSIONS:** Although the studies were conducted before statin therapy become standard care, and mostly in patients in secondary prevention, with various dosages of nicotinic acid 1-3g/day, this meta-analysis found positive effects of niacin alone or in combination on all cardiovascular events and on atherosclerosis evolution.

Question: Therapy

Study Type: Meta-analysis

Learner Level: intermediate

Notes: The title sounds impressive but there are all sorts of problems with it --- too broad a question, pooling with too much clinical & statistical heterogeneity, and weird use of OR. I would say this is a more difficult paper to use in groups who already understand the basics of what to look for.

Burgio, K. L., et al. (1998). "Behavioral vs drug treatment for urge urinary incontinence in older women: a randomized controlled trial." *JAMA* 280(23): 1995-2000.

CONTEXT: Urinary incontinence is a common condition caused by many factors with several treatment options. **OBJECTIVE:** To compare the effectiveness of biofeedback-assisted behavioral treatment with drug treatment and a placebo control condition for the treatment of urge and mixed urinary incontinence in older community-dwelling women. **DESIGN:** Randomized placebo-controlled trial conducted from 1989 to 1995. **SETTING:** University-based outpatient geriatric medicine clinic. **PATIENTS:** A volunteer sample of 197 women aged 55 to 92 years with urge urinary incontinence or mixed incontinence with urge as the predominant pattern. Subjects had to have urodynamic evidence of bladder dysfunction, be ambulatory, and not have dementia. **INTERVENTION:** Subjects were randomized to 4 sessions (8 weeks) of biofeedback-assisted behavioral treatment, drug treatment (with oxybutynin chloride, possible range of doses, 2.5 mg daily to 5.0 mg 3 times daily), or a placebo control condition. **MAIN OUTCOME MEASURES:** Reduction in the frequency of incontinent episodes as determined by bladder diaries, and patients' perceptions of improvement and their comfort and satisfaction with treatment. **RESULTS:** For all 3 treatment groups, reduction of incontinence was most pronounced early in treatment and progressed more gradually thereafter. Behavioral treatment, which yielded a mean 80.7% reduction of incontinence episodes, was significantly more effective than drug treatment (mean 68.5% reduction; $P=.04$) and both were more effective than the placebo control condition (mean 39.4% reduction; $P<.001$ and $P=.009$, respectively). Patient-perceived improvement was greatest for behavioral treatment (74.1% "much better" vs 50.9% and 26.9% for drug treatment and placebo, respectively). Only 14.0% of patients receiving behavioral treatment wanted to change to another treatment vs 75.5% in each of the other groups. **CONCLUSION:** Behavioral treatment is a safe and effective conservative intervention that should be made more readily available to patients as a first-line treatment for urge and mixed incontinence.

Question: Therapy

Study Type: RCT

Learner Level: Beginner

Notes: Good, well done RCT with clear methods including a patient flow diagram (fig. 2) which can be used for teaching; Some tricky thinking for the NNT calculations (hint: look at figure 3 to help you out...) Good discussion points: Can discuss 'stratification'; Can discuss intention to treat; Interesting spectrum of outcomes

Buse, J. B., et al. (2011). "Use of twice-daily exenatide in Basal insulin-treated patients with type 2 diabetes: a randomized, controlled trial." *Ann Intern Med* 154(2): 103-112.

BACKGROUND: Insulin replacement in diabetes often requires prandial intervention to reach hemoglobin A(c) (HbA(c)) targets. **OBJECTIVE:** To test whether twice-daily exenatide injections reduce HbA(c) levels more than placebo in people receiving insulin glargine. **DESIGN:** Parallel, randomized, placebo-controlled trial, blocked and stratified by HbA(c) level at site, performed from October 2008 to January 2010. Participants, investigators, and personnel conducting the study were masked to treatment assignments. (ClinicalTrials.gov registration number: NCT00765817) **SETTING:** 59 centers in 5 countries. **PATIENTS:** Adults with type 2 diabetes and an HbA(c) level of 7.1% to 10.5% who were receiving insulin glargine alone or in combination with metformin or pioglitazone (or both agents). **INTERVENTION:** Assignment by a centralized, computer-generated, random-sequence interactive voice-response system to exenatide, 10 microg twice daily, or placebo for 30 weeks. **MEASUREMENTS:** The primary outcome was change in HbA(c) level. Secondary outcomes included the percentage of participants with HbA(c) values of 7.0% or less and 6.5% or less, 7-point self-monitored glucose profiles, body weight, waist circumference, insulin dose, hypoglycemia, and adverse events. **RESULTS:** 112 of 138 exenatide recipients and 101 of 123 placebo recipients completed the study. The HbA(c) level decreased by 1.74% with exenatide and 1.04% with placebo (between-group difference, -0.69% [95% CI, -0.93% to -0.46%]; $P < 0.001$). Weight decreased by 1.8 kg with exenatide and increased by 1.0 kg with placebo (between-group difference, -2.7 kg [CI, -3.7 to -1.7]). Average increases in insulin dosage with exenatide and placebo were 13 U/d and 20 U/d. The estimated rate of minor hypoglycemia was similar between groups. Thirteen exenatide recipients and 1 placebo recipient discontinued the study because of adverse events ($P < 0.010$); rates of nausea (41% vs. 8%), diarrhea (18% vs. 8%), vomiting (18% vs. 4%), headache (14% vs. 4%), and constipation (10% vs. 2%) were higher with exenatide than with placebo. **LIMITATIONS:** The study was of short duration. There were slight imbalances between groups at baseline in terms of sex, use of concomitant glucose-lowering medications, and HbA(c) levels, and more exenatide recipients than placebo recipients withdrew because of adverse events. **CONCLUSION:** Adding twice-daily exenatide injections improved glycemic control without increased hypoglycemia or weight gain in participants with uncontrolled type 2 diabetes who were receiving insulin glargine treatment. Adverse events of exenatide included nausea, diarrhea, vomiting, headache, and constipation. **PRIMARY FUNDING SOURCE:** Alliance of Eli Lilly and Company and Amylin Pharmaceuticals.

Question: Therapy

Study Type: RCT

Learner Level: Intermediate

Notes: Nicely reported methods and a fairly straightforward RCT, however good example of article with a lot of spin / industry bias including faulty comparator. The paper seems to be written to obtain FDA approval, rather than to answer a patient-driven scientific question. Teaching points include good example of why we should read the methods / results and draw our own conclusions.

Carlberg, B., et al. (2004). "Atenolol in hypertension: is it a wise choice?" *Lancet* 364(9446): 1684-1689.

BACKGROUND: Atenolol is one of the most widely used beta blockers clinically, and has often been used as a reference drug in randomised controlled trials of hypertension. However, questions have been raised about atenolol as the best reference drug for comparisons with other antihypertensives. Thus, our aim was to systematically review the effect of atenolol on cardiovascular morbidity and mortality in hypertensive patients. **METHODS:** Reports were identified through searches of The Cochrane Library, MEDLINE, relevant textbooks, and by personal communication with established researchers in hypertension. Randomised controlled trials that assessed the effect of atenolol on cardiovascular morbidity or mortality in patients with primary hypertension were included. **FINDINGS:** We identified four studies that compared atenolol with placebo or no treatment, and five that compared atenolol with other antihypertensive drugs. Despite major differences in blood pressure lowering, there were no outcome differences between atenolol and placebo in the four studies, comprising 6825 patients, who were followed up for a mean of 4.6 years on all-cause mortality (relative risk 1.01 [95% CI 0.89-1.15]), cardiovascular mortality (0.99 [0.83-1.18]), or myocardial infarction (0.99 [0.83-1.19]). The risk of stroke, however, tended to be lower in the atenolol than in the placebo group (0.85 [0.72-1.01]). When atenolol was compared with other antihypertensives, there were no major differences in blood pressure lowering between the treatment arms. Our meta-analysis showed a significantly higher mortality (1.13 [1.02-1.25]) with atenolol treatment than with other active treatment, in the five studies comprising 17671 patients who were followed up for a mean of 4.6 years. Moreover, cardiovascular mortality also tended to be higher with atenolol treatment than with other antihypertensive treatment. Stroke was also more frequent with atenolol treatment. **INTERPRETATION:** Our results cast doubts on atenolol as a suitable drug for hypertensive patients. Moreover, they challenge the use of atenolol as a reference drug in outcome trials in hypertension.

Question: Therapy

Study Type: Meta-analysis

Learner Level: Beginner

Notes: This is a nice straightforward meta-analysis that shows how a commonly used drug may not end up providing as much clinical benefit as originally thought. Can also be used to discuss how a surrogate end-point (blood pressure) does not always reflect the true impact on clinical events.

Chan, F. K., et al. (2010). "Celecoxib versus omeprazole and diclofenac in patients with osteoarthritis and rheumatoid arthritis (CONDOR): a randomised trial." *Lancet* 376(9736): 173-179.

BACKGROUND: Cyclo-oxygenase (COX)-2-selective non-steroidal anti-inflammatory drugs (NSAIDs) and non-selective NSAIDs plus a proton-pump inhibitor (PPI) have similar upper gastrointestinal outcomes, but risk of clinical outcomes across the entire gastrointestinal tract might be lower with selective drugs than with non-selective drugs. We aimed to compare risk of gastrointestinal events associated with celecoxib versus diclofenac slow release plus omeprazole. **METHODS:** We undertook a 6-month, double-blind, randomised trial in patients with osteoarthritis or rheumatoid arthritis at increased gastrointestinal risk at 196 centres in 32 countries or territories. Patients tested negative for *Helicobacter pylori* and were aged 60 years and older or 18 years and older with previous gastroduodenal ulceration. We used a computer-generated randomisation schedule to assign patients in a 1:1 ratio to receive celecoxib 200 mg twice a day or diclofenac slow release 75 mg twice a day plus omeprazole 20 mg once a day. Patients and investigators were masked to treatment allocation. The primary endpoint was a composite of clinically significant upper or lower gastrointestinal events adjudicated by an independent committee. Analysis was by intention to treat. This trial is registered with ClinicalTrials.gov, number NCT00141102. **FINDINGS:** 4484 patients were randomly allocated to treatment (2238 celecoxib; 2246 diclofenac plus omeprazole) and were included in intention-to-treat analyses. 20 (0.9%) patients receiving celecoxib and 81 (3.8%) receiving diclofenac plus omeprazole met criteria for the primary endpoint (hazard ratio 4.3, 95% CI 2.6-7.0; $p < 0.0001$). 114 (6%) patients taking celecoxib versus 167 (8%) taking diclofenac plus omeprazole withdrew early because of gastrointestinal adverse events ($p = 0.0006$). **INTERPRETATION:** Risk of clinical outcomes throughout the gastrointestinal tract was lower in patients treated with a COX-2-selective NSAID than in those receiving a non-selective NSAID plus a PPI. These findings should encourage review of approaches to reduce risk of NSAID treatment. **FUNDING:** Pfizer Inc.

Question: Therapy

Study Type: RCT

Learner Level: Beginner - Intermediate

Notes: A therapy article which can be useful for basic therapy validity and calculating ARR, RRR, NNT. Also raises the issue of the composite outcome, and which components of the composite drive the results. Can be beginner to intermediate.

Chastre, J., et al. (2003). "Comparison of 8 vs 15 days of antibiotic therapy for ventilator-associated pneumonia in adults: a randomized trial." *JAMA* 290(19): 2588-2598.

CONTEXT: The optimal duration of antimicrobial treatment for ventilator-associated pneumonia (VAP) is unknown. Shortening the length of treatment may help to contain the emergence of multiresistant bacteria in the intensive care unit (ICU). **OBJECTIVE:** To determine whether 8 days is as effective as 15 days of antibiotic treatment of patients with microbiologically proven VAP. **DESIGN, SETTING, AND PARTICIPANTS:** Prospective, randomized, double-blind (until day 8) clinical trial conducted in 51 French ICUs. A total of 401 patients diagnosed as having developed VAP by quantitative culture results of bronchoscopic specimens and who had received initial appropriate empirical antimicrobial therapy were enrolled between May 1999 and June 2002. **INTERVENTION:** A total of 197 patients were randomly assigned to receive 8 days and 204 to receive 15 days of therapy with an antibiotic regimen selected by the treating physician. **MAIN OUTCOME MEASURES:** Primary outcome measures—death from any cause, microbiologically documented pulmonary infection recurrence, and antibiotic-free days—were assessed 28 days after VAP onset and analyzed on an intent-to-treat basis. **RESULTS:** Compared with patients treated for 15 days, those treated for 8 days had neither excess mortality (18.8% vs 17.2%; difference, 1.6%; 90% confidence interval [CI], -3.7% to 6.9%) nor more recurrent infections (28.9% vs 26.0%; difference, 2.9%; 90% CI, -3.2% to 9.1%), but they had more mean (SD) antibiotic-free days (13.1 [7.4] vs 8.7 [5.2] days, $P < .001$). The number of mechanical ventilation-free days, the number of organ failure-free days, the length of ICU stay, and mortality rates on day 60 for the 2 groups did not differ. Although patients with VAP caused by nonfermenting gram-negative bacilli, including *Pseudomonas aeruginosa*, did not have more unfavorable outcomes when antimicrobial therapy lasted only 8 days, they did have a higher pulmonary infection-recurrence rate compared with those receiving 15 days of treatment (40.6% vs 25.4%; difference, 15.2%, 90% CI, 3.9%-26.6%). Among patients who developed recurrent infections, multiresistant pathogens emerged less frequently in those who had received 8 days of antibiotics (42.1% vs 62.0% of pulmonary recurrences, $P = .04$). **CONCLUSIONS:** Among patients who had received appropriate initial empirical therapy, with the possible exception of those developing nonfermenting gram-negative bacillus infections, comparable clinical effectiveness against VAP was obtained with the 8- and 15-day treatment regimens. The 8-day group had less antibiotic use.

Question: Therapy

Study Type: RCT

Learner Level: Intermediate

Notes: Noninferiority Trial; Good methods for discussion of RCT including blinding, allocation concealment—the trial was blinded for the first 8 days of the study and then antibiotics continued or not continued by random assignment from days 9 through 15. Given the study question and patient population it is a nice study design. Not necessarily for beginners, but for more advanced learners this is a nice study about an important topic (duration of therapy). Can calculate NNH for superinfection / relapse from Table 4

Cramer, H., et al. (2013). "A systematic review and meta-analysis of yoga for low back pain." *Clinical Journal of Pain* 29(5): 450-460.

OBJECTIVES: To systematically review and meta-analyze the effectiveness of yoga for low back pain. **METHODS:** MEDLINE, the Cochrane Library, EMBASE, CAMBASE, and PsycINFO, were screened through January 2012. Randomized controlled trials comparing yoga to control conditions in patients with low back pain were included. Two authors independently assessed risk of bias using the risk of bias tool recommended by the Cochrane Back Review Group. Main outcome measures were pain, back-specific disability, generic disability, health-related quality of life, and global improvement. For each outcome, standardized mean differences (SMD) and 95% confidence intervals (CI) were calculated. **RESULTS:** Ten randomized controlled trials with a total of 967 chronic low back pain patients were included. Eight studies had low risk of bias. There was strong evidence for short-term effects on pain (SMD=-0.48; 95% CI, -0.65 to -0.31; P<0.01), back-specific disability (SMD=-0.59; 95% CI, -0.87 to -0.30; P<0.01), and global improvement (risk ratio=3.27; 95% CI, 1.89-5.66; P<0.01). There was strong evidence for a long-term effect on pain (SMD=-0.33; 95% CI, -0.59 to -0.07; P=0.01) and moderate evidence for a long-term effect on back-specific disability (SMD=-0.35; 95% CI, -0.55 to -0.15; P<0.01). There was no evidence for either short-term or long-term effects on health-related quality of life. Yoga was not associated with serious adverse events. **DISCUSSION:** This systematic review found strong evidence for short-term effectiveness and moderate evidence for long-term effectiveness of yoga for chronic low back pain in the most important patient-centered outcomes. Yoga can be recommended as an additional therapy to chronic low back pain patients.

Question: Therapy

Study Type: Meta-Analysis

Learner Level: Beginner

Notes: Beginner to intermediate level, simple, well-done meta-analysis of a common question, easy to teach I2, instructive Forest plots.

Cryer, P. E., et al. (2009). "Evaluation and management of adult hypoglycemic disorders: an Endocrine Society Clinical Practice Guideline." *J Clin Endocrinol Metab* 94(3): 709-728.

OBJECTIVE: The aim is to provide guidelines for the evaluation and management of adults with hypoglycemic disorders, including those with diabetes mellitus. **EVIDENCE:** Using the recommendations of the Grading of Recommendations, Assessment, Development and Evaluation (GRADE) system, the quality of evidence is graded very low (plus sign in circle ooo), low (plus sign in circle plus sign in circle oo), moderate (plus sign in circle plus sign in circle plus sign in circle o), or high (plus sign in circle plus sign in circle plus sign in circle). **CONCLUSIONS:** We recommend evaluation and management of hypoglycemia only in patients in whom Whipple's triad--symptoms, signs, or both consistent with hypoglycemia, a low plasma glucose concentration, and resolution of those symptoms or signs after the plasma glucose concentration is raised--is documented. In patients with hypoglycemia without diabetes mellitus, we recommend the following strategy. First, pursue clinical clues to potential hypoglycemic etiologies--drugs, critical illnesses, hormone deficiencies, nonislet cell tumors. In the absence of these causes, the differential diagnosis narrows to accidental, surreptitious, or even malicious hypoglycemia or endogenous hyperinsulinism. In patients suspected of having endogenous hyperinsulinism, measure plasma glucose, insulin, C-peptide, proinsulin, beta-hydroxybutyrate, and circulating oral hypoglycemic agents during an episode of hypoglycemia and measure insulin antibodies. Insulin or insulin secretagogue treatment of diabetes mellitus is the most common cause of hypoglycemia. We recommend the practice of hypoglycemia risk factor reduction--addressing the issue of hypoglycemia, applying the principles of intensive glycemic therapy, and considering both the conventional risk factors and those indicative of compromised defenses against falling plasma glucose concentrations--in persons with diabetes.

Question: Therapy

Study Type: Guideline

Learner Level: Intermediate

Notes: Hypoglycemia guideline: GRADE is emerging as the most coherent and comprehensive approach to guideline development, healthcare providers need to become fluent in guideline appraisal and interpretation using GRADE-developed CPGs.

Damoiseaux, R. A., et al. (2000). "Primary care based randomised, double blind trial of amoxicillin versus placebo for acute otitis media in children aged under 2 years." *BMJ* 320(7231): 350-354.

OBJECTIVE: To determine the effect of antibiotic treatment for acute otitis media in children between 6 months and 2 years of age. **DESIGN:** Practice based, double blind, randomised, placebo controlled trial. **SETTING:** 53 general practices in the Netherlands. **SUBJECTS:** 240 children aged 6 months to 2 years with the diagnosis of acute otitis media. **INTERVENTION:** Amoxicillin 40 mg/kg/day in three doses. **MAIN OUTCOME MEASURES:** Persistent symptoms at day four and duration of fever and pain or crying, or both. Otoscopy at days four and 11, tympanometry at six weeks, and use of analgesic. **RESULTS:** Persistent symptoms at day four were less common in the amoxicillin group (risk difference 13%; 95% confidence interval 1% to 25%). The median duration of fever was two days in the amoxicillin group versus three in the placebo group (P=0.004). No significant difference was observed in duration of pain or crying, but analgesic consumption was higher in the placebo group during the first 10 days (4.1 v 2.3 doses, P=0.004). In addition, no otoscopic differences were observed at days four and 11, and tympanometric findings at six weeks were similar in both groups. **CONCLUSIONS:** Seven to eight children aged 6 to 24 months with acute otitis media needed to be treated with antibiotics to improve symptomatic outcome at day four in one child. This modest effect does not justify prescription of antibiotics at the first visit, provided close surveillance can be guaranteed.

Question: Therapy

Study Type: RCT

Learner Level: Beginner

Notes: Fairly simple design with clear participant flow diagrams. The inclusion criteria and diagnostic criteria provide an opportunity to discuss how patient selection may bias the results in this study. Good discussion points: Applicable to parents and pediatricians alike. May provoke some interesting discussion about outcomes of interest and the conclusions that the authors draw from the data.

De Berardis, G., et al. (2009). "Aspirin for primary prevention of cardiovascular events in people with diabetes: meta-analysis of randomised controlled trials." BMJ 339: b4531.

OBJECTIVE: To evaluate the benefits and harms of low dose aspirin in people with diabetes and no cardiovascular disease. **DESIGN:** Meta-analysis of randomised controlled trials. **DATA SOURCES:** Medline (1966-November 2008), the Cochrane central register of controlled trials (Cochrane Library 2008;issue 4), and reference lists of retrieved articles. Review methods Randomised trials of aspirin compared with placebo or no aspirin in people with diabetes and no pre-existing cardiovascular disease were eligible for inclusion. Data on major cardiovascular events (death from cardiovascular causes, non-fatal myocardial infarction, non-fatal stroke, and all cause mortality) were extracted and pooled with a random effect model. Results are reported as relative risks with 95% confidence intervals. **RESULTS:** Of 157 studies in the literature searches, six were eligible (10 117 participants). When aspirin was compared with placebo there was no statistically significant reduction in the risk of major cardiovascular events (five studies, 9584 participants; relative risk 0.90, 95% confidence interval 0.81 to 1.00), cardiovascular mortality (four studies, n=8557, 0.94; 0.72 to 1.23), or all cause mortality (four studies, n=8557; 0.93, 0.82 to 1.05). Significant heterogeneity was found in the analysis for myocardial infarction (I(2)=62.2%; P=0.02) and stroke (I(2)=52.5%; P=0.08). Aspirin significantly reduced the risk of myocardial infarction in men (0.57, 0.34 to 0.94) but not in women (1.08, 0.71 to 1.65; P for interaction=0.056). Evidence relating to harms was inconsistent. **CONCLUSIONS:** A clear benefit of aspirin in the primary prevention of major cardiovascular events in people with diabetes remains unproved. Sex may be an important effect modifier. Toxicity is to be explored further.

Question: Therapy

Study Type: Meta-analysis

Learner Level: Beginner - Intermediate

Notes: Beginner / Intermediate meta-analysis of aspirin for primary prevention in diabetics; Large number of patients, nice forest plots; Very good for teaching about heterogeneity and I-square

De, B. K., et al. (2009). "Pentoxifylline versus prednisolone for severe alcoholic hepatitis: a randomized controlled trial." World J Gastroenterol 15(13): 1613-1619.

AIM: To compare the efficacy of pentoxifylline and prednisolone in the treatment of severe alcoholic hepatitis, and to evaluate the role of different liver function scores in predicting prognosis. **METHODS:** Sixty-eight patients with severe alcoholic hepatitis (Maddrey score > or = 32) received pentoxifylline (n = 34, group I) or prednisolone (n = 34, group II) for 28 d in a randomized double-blind controlled study, and subsequently in an open study (with a tapering dose of prednisolone) for a total of 3 mo, and were followed up over a period of 12 mo. **RESULTS:** Twelve patients in group II died at the end of 3 mo in contrast to five patients in group I. The probability of dying at the end of 3 mo was higher in group II as compared to group I (35.29% vs 14.71%, P = 0.04; log rank test). Six patients in group II developed hepatorenal syndrome as compared to none in group I. Pentoxifylline was associated with a significantly lower model for end-stage liver disease (MELD) score at the end of 28 d of therapy (15.53 +/- 3.63 vs 17.78 +/- 4.56, P = 0.04). Higher baseline Maddrey score was associated with increased mortality. **CONCLUSION:** Reduced mortality, improved risk-benefit profile and renoprotective effects of pentoxifylline compared with prednisolone suggest that pentoxifylline is superior to prednisolone for treatment of severe alcoholic hepatitis.

Question: Therapy

Study Type: RCT

Learner Level: Intermediate

Notes: Good to teach NNT and ARR

de Ferranti, S. D., et al. (1998). "Are amoxicillin and folate inhibitors as effective as other antibiotics for acute sinusitis? A meta-analysis." BMJ 317(7159): 632-637.

OBJECTIVES: To examine whether antibiotics are indicated in treating uncomplicated acute sinusitis and, if so, whether newer and more expensive antibiotics with broad spectra of antimicrobial activity are more effective than amoxicillin or folate inhibitors. **DESIGN:** Meta-analysis of randomised trials. **SETTING:** Outpatient clinics. **SUBJECTS:** 2717 patients with acute sinusitis or acute exacerbation of chronic sinusitis from 27 trials. **INTERVENTIONS:** Any antibiotic versus placebo; amoxicillin or folate inhibitors versus newer, more expensive antibiotics. **MAIN OUTCOME MEASUREMENTS:** Clinical failures and cures. **RESULTS:** Compared with placebo,

antibiotics decreased the incidence of clinical failures by half (risk ratio 0.54 (95% confidence interval 0.37 to 0.79)). Risk of clinical failure among 1553 randomised patients was not meaningfully decreased with more expensive antibiotics as compared with amoxicillin (risk ratio 0.86 (0.62 to 1.19); risk difference 0.9 fewer failures per 100 patients (1.4 more failures to 3.1 fewer failures per 100 patients)). The results were similar for other antibiotics versus folate inhibitors (risk ratio 1.01 (0.52 to 1.97)), but data were sparse (n=410) and of low quality. CONCLUSIONS: Amoxicillin and folate inhibitors are essentially as effective as more expensive antibiotics for the initial treatment of uncomplicated acute sinusitis. Small differences in efficacy may exist, but are unlikely to be clinically important.

Question: Therapy

Study Type: Systematic Review

Learner Level: Beginner / Intermediate

Notes: There are nice tables and figures that are classic for systematic review and meta-analysis. Good discussion points: Good discussion of both a positive finding (antibiotics vs. placebo) and also negative finding (cheap vs. expensive antibiotics). Also good to discuss cost implications.

Dentali, F., et al. (2007). "Meta-analysis: anticoagulant prophylaxis to prevent symptomatic venous thromboembolism in hospitalized medical patients." *Ann Intern Med* 146(4): 278-288.

BACKGROUND: Underutilization of anticoagulant prophylaxis may be due to lack of evidence that prophylaxis prevents clinically important outcomes in hospitalized medical patients at risk for venous thromboembolism. PURPOSE: To assess the effects of anticoagulant prophylaxis in reducing clinically important outcomes in hospitalized medical patients. DATA SOURCES: MEDLINE, EMBASE, and Cochrane databases were searched to September 2006 without language restrictions. STUDY SELECTION: Randomized trials comparing anticoagulant prophylaxis with no treatment in hospitalized medical patients. DATA EXTRACTION: Any symptomatic pulmonary embolism (PE), fatal PE, symptomatic deep venous thrombosis, all-cause mortality, and major bleeding. Pooled relative risks and associated 95% CIs were calculated. For treatment effects that were statistically significant, the authors determined the absolute risk reduction and the number needed to treat for benefit (NNT(B)) to prevent an outcome. DATA SYNTHESIS: 9 studies (n = 19 958) were included. During anticoagulant prophylaxis, patients had significant reductions in any PE (relative risk, 0.43 [CI, 0.26 to 0.71]; absolute risk reduction, 0.29%; NNT(B), 345) and fatal PE (relative risk, 0.38 [CI, 0.21 to 0.69]; absolute risk reduction, 0.25%; NNT(B), 400), a nonsignificant reduction in symptomatic deep venous thrombosis (relative risk, 0.47 [CI, 0.22 to 1.00]), and a nonsignificant increase in major bleeding (relative risk, 1.32 [CI, 0.73 to 2.37]). Anticoagulant prophylaxis had no effect on all-cause mortality (relative risk, 0.97 [CI, 0.79 to 1.19]). LIMITATIONS: 2 of 9 included studies were not double-blind. CONCLUSIONS: Anticoagulant prophylaxis is effective in preventing symptomatic venous thromboembolism during anticoagulant prophylaxis in at-risk hospitalized medical patients. Additional research is needed to determine the risk for venous thromboembolism in these patients after prophylaxis has been stopped.

Question: Therapy

Study Type: Meta-analysis

Learner Level: Intermediate

Notes: Paper of importance to internists, hospitalists, house officers and anyone who takes care of patients in the hospital. Clear methodology with excellent forest plots for teaching. Can use to illustrate a paper where results do not display any heterogeneity.

Devereaux, P. J., et al. (2005). "How strong is the evidence for the use of perioperative beta blockers in non-cardiac surgery? Systematic review and meta-analysis of randomised controlled trials." *BMJ* 331(7512): 313-321.

OBJECTIVE: To determine the effect of perioperative beta blocker treatment in patients having non-cardiac surgery. DESIGN: Systematic review and meta-analysis. DATA SOURCES: Seven search strategies, including searching two bibliographic databases and hand searching seven medical journals. STUDY SELECTION AND OUTCOMES: We included randomised controlled trials that evaluated beta blocker treatment in patients having non-cardiac surgery. Perioperative outcomes within 30 days of surgery included total mortality, cardiovascular mortality, non-fatal myocardial infarction, non-fatal cardiac arrest, non-fatal stroke, congestive heart failure, hypotension needing treatment, bradycardia needing treatment, and bronchospasm. RESULTS: Twenty two trials that randomised a total of 2437 patients met the eligibility criteria. Perioperative beta blockers did not show any statistically significant beneficial effects on any of the individual outcomes and the only nominally statistically significant beneficial relative risk was 0.44 (95% confidence interval 0.20 to 0.97, 99% confidence interval 0.16 to 1.24) for the composite outcome of cardiovascular mortality, non-fatal myocardial infarction, and non-fatal cardiac arrest. Methods adapted from formal interim monitoring boundaries applied to cumulative meta-analysis showed that the evidence failed, by a considerable degree, to meet standards for forgoing additional studies. The individual safety outcomes in patients treated with perioperative beta blockers showed a relative risk for bradycardia needing treatment of 2.27 (95% CI 1.53 to 3.36, 99% CI 1.36 to 3.80) and a nominally statistically significant relative risk for hypotension needing treatment of 1.27 (95% CI 1.04 to 1.56, 99% CI 0.97 to 1.66). CONCLUSION: The evidence that perioperative beta blockers reduce major cardiovascular events is encouraging but too unreliable to allow definitive conclusions to be drawn.

Question: Therapy

Study Type: Meta-analysis

Learner Level: Intermediate

Notes: Extremely rigorous methods clearly reported make this a great teaching paper. In addition, results are counter to current practice of broadly applying b-blockers in the perioperative period to patient undergoing non-cardiac surgery. Results are good for teaching about heterogeneity. The section in the results section on exploring heterogeneity can be very instructive.

DiazGranados, C. A., et al. (2014). "Efficacy of high-dose versus standard-dose influenza vaccine in older adults." *N Engl J Med* 371(7): 635-645.

BACKGROUND: As compared with a standard-dose vaccine, a high-dose, trivalent, inactivated influenza vaccine (IIV3-HD) improves antibody responses to influenza among adults 65 years of age or older. This study evaluated whether IIV3-HD also improves protection against laboratory-confirmed influenza illness. **METHODS:** We conducted a phase IIIb-IV, multicenter, randomized, double-blind, active-controlled trial to compare IIV3-HD (60 mug of hemagglutinin per strain) with standard-dose trivalent, inactivated influenza vaccine (IIV3-SD [15 mug of hemagglutinin per strain]) in adults 65 years of age or older. Assessments of relative efficacy, effectiveness, safety (serious adverse events), and immunogenicity (hemagglutination-inhibition [HAI] titers) were performed during the 2011-2012 (year 1) and the 2012-2013 (year 2) northern-hemisphere influenza seasons. **RESULTS:** A total of 31,989 participants were enrolled from 126 research centers in the United States and Canada (15,991 were randomly assigned to receive IIV3-HD, and 15,998 to receive IIV3-SD). In the intention-to-treat analysis, 228 participants in the IIV3-HD group (1.4%) and 301 participants in the IIV3-SD group (1.9%) had laboratory-confirmed influenza caused by any viral type or subtype associated with a protocol-defined influenza-like illness (relative efficacy, 24.2%; 95% confidence interval [CI], 9.7 to 36.5). At least one serious adverse event during the safety surveillance period was reported by 1323 (8.3%) of the participants in the IIV3-HD group, as compared with 1442 (9.0%) of the participants in the IIV3-SD group (relative risk, 0.92; 95% CI, 0.85 to 0.99). After vaccination, HAI titers and seroprotection rates (the percentage of participants with HAI titers \geq 1:40) were significantly higher in the IIV3-HD group. **Conclusions:** Among persons 65 years of age or older, IIV3-HD induced significantly higher antibody responses and provided better protection against laboratory-confirmed influenza illness than did IIV3-SD. (Funded by Sanofi Pasteur; ClinicalTrials.gov number, NCT01427309.)

Question: Therapy

Study Type: RCT

Learner Level: Intermediate

Notes: paper is a good teaching paper for helping learners identify industry bias. The study design is strong in terms of randomization, allocation concealment, blinding, ITT. However, the results are not meaningful in the setting of non-clinically important outcomes. All but 2 of the authors have received funds from the makers of the vaccine or are direct employees / stock holders. This article can be used for focus on why critical reading is necessary to determine whether evidence is generalizable for implementation.

Dorian, P., et al. (2002). "Amiodarone as compared with lidocaine for shock-resistant ventricular fibrillation." *N Engl J Med* 346(12): 884-890.

BACKGROUND: Lidocaine has been the initial antiarrhythmic drug treatment recommended for patients with ventricular fibrillation that is resistant to conversion by defibrillator shocks. We performed a randomized trial comparing intravenous lidocaine with intravenous amiodarone as an adjunct to defibrillation in victims of out-of-hospital cardiac arrest. **METHODS:** Patients were enrolled if they had out-of-hospital ventricular fibrillation resistant to three shocks, intravenous epinephrine, and a further shock; or if they had recurrent ventricular fibrillation after initially successful defibrillation. They were randomly assigned in a double-blind manner to receive intravenous amiodarone plus lidocaine placebo or intravenous lidocaine plus amiodarone placebo. The primary end point was the proportion of patients who survived to be admitted to the hospital. **RESULTS:** In total, 347 patients (mean \pm SD age, 67 \pm 14 years) were enrolled. The mean interval between the time at which paramedics were dispatched to the scene of the cardiac arrest and the time of their arrival was 7 \pm 3 minutes, and the mean interval from dispatch to drug administration was 25 \pm 8 minutes. After treatment with amiodarone, 22.8 percent of 180 patients survived to hospital admission, as compared with 12.0 percent of 167 patients treated with lidocaine ($P=0.009$; odds ratio, 2.17; 95 percent confidence interval, 1.21 to 3.83). Among patients for whom the time from dispatch to the administration of the drug was equal to or less than the median time (24 minutes), 27.7 percent of those given amiodarone and 15.3 percent of those given lidocaine survived to hospital admission ($P=0.05$). **CONCLUSIONS:** As compared with lidocaine, amiodarone leads to substantially higher rates of survival to hospital admission in patients with shock-resistant out-of-hospital ventricular fibrillation.

Question: Therapy

Study Type: RCT

Learner Level: Beginner / Intermediate

Notes: Easy to do in a 1 hour session. Valid with important findings Teaching points: Challenges of doing RCT in real life = good for discussions about efficacy vs effectiveness. Allows for good discussion on NNT. Allows for good discussion on intention to treat

Drummond, M. B., et al. (2008). "Inhaled corticosteroids in patients with stable chronic obstructive pulmonary disease: a systematic review and meta-analysis." *JAMA* 300(20): 2407-2416.

CONTEXT: Recent studies of inhaled corticosteroid (ICS) therapy for managing stable chronic obstructive pulmonary disease (COPD) have yielded conflicting results regarding survival and risk of adverse events. OBJECTIVE: To systematically review and quantitatively synthesize the effects of ICS therapy on mortality and adverse events in patients with stable COPD. DATA SOURCES: Search of MEDLINE, CENTRAL, EMBASE, CINAHL, Web of Science, and PsychInfo through February 9, 2008. STUDY SELECTION: Eligible studies were double-blind, randomized controlled trials comparing ICS therapy for 6 or more months with nonsteroid inhaled therapy in patients with COPD. DATA EXTRACTION: Two authors independently abstracted data including study characteristics, all-cause mortality, pneumonia, and bone fractures. The I(2) statistic was used to assess heterogeneity. Study-level data were pooled using a random-effects model (when I(2) > or = 50%) or a fixed-effects model (when I(2) < 50%). For the primary outcome of all-cause mortality at 1 year, our meta-analysis was powered to detect a 1.0% absolute difference in mortality, assuming a 2-sided alpha of .05 and power of 0.80. RESULTS: Eleven eligible randomized controlled trials (14,426 participants) were included. In trials with mortality data, no difference was observed in 1-year all-cause mortality (128 deaths among 4636 patients in the treatment group and 148 deaths among 4597 patients in the control group; relative risk [RR], 0.86; 95% confidence interval [CI], 0.68-1.09; P = .20; I(2) = 0%). In the trials with data on pneumonia, ICS therapy was associated with a significantly higher incidence of pneumonia (777 cases among 5405 patients in the treatment group and 561 cases among 5371 patients in the control group; RR, 1.34; 95% CI, 1.03-1.75; P = .03; I(2) = 72%). Subgroup analyses indicated an increased risk of pneumonia in the following subgroups: highest ICS dose (RR, 1.46; 95% CI, 1.10-1.92; P = .008; I(2) = 78%), shorter duration of ICS use (RR, 2.12; 95% CI, 1.47-3.05; P < .001; I(2) = 0%), lowest baseline forced expiratory volume in the first second of expiration (RR, 1.90; 95% CI, 1.26-2.85; P = .002; I(2) = 0%), and combined ICS and bronchodilator therapy (RR, 1.57; 95% CI, 1.35-1.82; P < .001; I(2) = 24%). CONCLUSIONS: Among patients with COPD, ICS therapy does not affect 1-year all-cause mortality. ICS therapy is associated with a higher risk of pneumonia. Future studies should determine whether specific subsets of patients with COPD benefit from ICS therapy.

Question: Therapy

Study Type: Meta-analysis

Learner Level: Intermediate to Advanced

Notes: Clearly reported meta-analysis on a somewhat controversial topic (inhaled steroids for COPD) with a controversial result (no change in mortality but increased pneumonia in treated group). This is also a good paper for discussing heterogeneity: mortality outcome in figure 2 (I²=0), but pneumonia risk has significant heterogeneity indicating that it might not be okay to combine (I²=72%). This paper should be reserved for more advanced learners who are ready to grapple with more difficult issues. There is an ACP Journal Club summary of this paper, but remarkably they don't mention the heterogeneity issues.

Duckworth, W., et al. (2009). "Glucose control and vascular complications in veterans with type 2 diabetes." *N Engl J Med* 360(2): 129-139.

BACKGROUND: The effects of intensive glucose control on cardiovascular events in patients with long-standing type 2 diabetes mellitus remain uncertain. METHODS: We randomly assigned 1791 military veterans (mean age, 60.4 years) who had a suboptimal response to therapy for type 2 diabetes to receive either intensive or standard glucose control. Other cardiovascular risk factors were treated uniformly. The mean number of years since the diagnosis of diabetes was 11.5, and 40% of the patients had already had a cardiovascular event. The goal in the intensive-therapy group was an absolute reduction of 1.5 percentage points in the glycated hemoglobin level, as compared with the standard-therapy group. The primary outcome was the time from randomization to the first occurrence of a major cardiovascular event, a composite of myocardial infarction, stroke, death from cardiovascular causes, congestive heart failure, surgery for vascular disease, inoperable coronary disease, and amputation for ischemic gangrene. RESULTS: The median follow-up was 5.6 years. Median glycated hemoglobin levels were 8.4% in the standard-therapy group and 6.9% in the intensive-therapy group. The primary outcome occurred in 264 patients in the standard-therapy group and 235 patients in the intensive-therapy group (hazard ratio in the intensive-therapy group, 0.88; 95% confidence interval [CI], 0.74 to 1.05; P=0.14). There was no significant difference between the two groups in any component of the primary outcome or in the rate of death from any cause (hazard ratio, 1.07; 95% CI, 0.81 to 1.42; P=0.62). No differences between the two groups were observed for microvascular complications. The rates of adverse events, predominantly hypoglycemia, were 17.6% in the standard-therapy group and 24.1% in the intensive-therapy group. CONCLUSIONS: Intensive glucose control in patients with poorly controlled type 2 diabetes had no significant effect on the rates of major cardiovascular events, death, or microvascular complications with the exception of progression of albuminuria (P = 0.01) [added]. (ClinicalTrials.gov number, NCT00032487.)

Question: Therapy

Study Type: RCT

Learner Level: Intermediate

Notes: Clearly reported methods that allow discussion of RCT validity criteria. This is one of several papers to come out within a 6 month period that does not support more aggressive glucose management. Good for teaching Kaplan Meier curves as well as calculating Number needed to harm (NNH). (see also the ACCORD study NEJM 2008; 358:2545-59.)

Duffy, L. M., et al. (1995). "Clean intermittent catheterization: safe, cost-effective bladder management for male residents of VA nursing homes." *J Am Geriatr Soc* 43(8): 865-870.

OBJECTIVES: To compare the safety and cost of clean versus sterile intermittent bladder catheterization in male nursing home residents. To provide evidence to support the hypothesis that intermittent catheterization is a valid, alternative method of bladder management in male residents of long-term care in whom urinary retention is a documented problem. DESIGN: Randomized clinical trial. SETTING: Three long-term care sites having predominantly male populations. PARTICIPANTS: Eighty male veterans, residents of

three long-term care facilities, ranging in age from 36 to 96 years with a mean age of 72. INTERVENTIONS: Standardized procedures for clean and sterile intermittent catheterization (IC) were implemented by staff nurses at each site. Patients were randomized into clean and sterile IC groups. Nursing time and catheterization equipment usage were recorded using bar code readers. Clinical data were collected from the medical chart. Treatment of urinary tract infection was prescribed by the medical personnel responsible for each individual resident. MEASUREMENTS: We compared the number of treatment episodes for symptomatic bacteriuria between groups randomized to receive either clean or sterile intermittent catheterization. Laboratory analysis of blood and urine was done on predetermined days. Control variables were research site and patient history of urinary tract infection within the last 6 months. A cost comparison of nursing time and equipment usage for the two catheterization techniques was also performed. RESULTS: No significant differences were found between clean and sterile groups with regard to number of treatment episodes, time to first infection, type of organism cultured, or cost of antibiotic treatment. The cost of sterile technique was considerably higher both in terms of nursing time and supplies. CONCLUSIONS: Findings from this study demonstrate that clean technique intermittent catheterization is a safe and cost-effective bladder management technique with male, nursing home residents, despite the frailty of this high risk population. An annual savings of approximately \$1460 per patient in nursing time and catheterization supplies could be anticipated if a patient were catheterized an average of four times per day substituting clean IC technique for sterile IC technique.

Question: Therapy

Study Type: RCT

Learner Level: Beginner / Intermediate

Notes: Straightforward RCT with clear methods; Equivalence trial, therefore can not calculate an NNT. Good discussion points: Great multidisciplinary paper; Can discuss 'stratification' (this is hidden in the methods sections, you may need to figure out where it shows up...) Might use this paper to discuss sample size see page 868); Good discussion of costs

Eidelman, R. S., et al. (2004). "Randomized trials of vitamin E in the treatment and prevention of cardiovascular disease." *Arch Intern Med* 164(14): 1552-1556.

BACKGROUND: Observational epidemiological studies consistently show that individuals who choose to take high amounts of vitamin E through diet or supplements experience cardiovascular benefits, for which basic research provides plausible mechanisms. However, because the size of the postulated benefit is small to moderate, the confounding inherent in observational studies is as great as the effect size. Before the availability of randomized evidence, about 1 in 4 adults was taking vitamin E supplements in the United States. **METHODS:** We conducted a computerized search of the English-language literature from 1990 to the present and found 7 large-scale randomized trials of the effectiveness vitamin E in the treatment and prevention of cardiovascular disease. Data were available on myocardial infarction, stroke, or cardiovascular death. **RESULTS:** Six of the 7 trials showed no significant effect of vitamin E on cardiovascular disease. In an overview, vitamin E had neither a statistically significant nor a clinically important effect on any important cardiovascular event (odds ratio [OR], 0.98; 95% confidence interval [CI], 0.94-1.03) or its components: nonfatal myocardial infarction (OR, 1.00; 95% CI, 0.92-1.09), nonfatal stroke (OR, 1.03; 95% CI, 0.93-1.14), or cardiovascular death (OR, 1.00; 95% CI, 0.94-1.05). **CONCLUSIONS:** The ORs and CIs provide strong support for a lack of statistically significant or clinically important effects of vitamin E on cardiovascular disease. The use of agents of proven lack of benefit, especially those easily available over the counter, may contribute to underuse of agents of proven benefit and failure to adopt healthy lifestyles.

Question: Therapy

Study Type: Systematic Review

Learner Level: Intermediate

Notes: Very poor quality systematic review. Give very little information on methods. Good Discussion Points: This is an excellent article to use in combination with the Shekelle article to highlight the differences in quality. This is a good example of inappropriate pooling of studies. Pooling combined antioxidants with vitamin E alone. Good example of how poor quality articles can be published in well regarded journals. Can be used to force learners to think about how they would have done the review.

Fernandez, J., et al. (2007). "Primary prophylaxis of spontaneous bacterial peritonitis delays hepatorenal syndrome and improves survival in cirrhosis." *Gastroenterology* 133(3): 818-824.

BACKGROUND & AIMS: Norfloxacin is highly effective in preventing spontaneous bacterial peritonitis recurrence in cirrhosis, but its role in the primary prevention of this complication is uncertain. **METHODS:** Patients with cirrhosis and low protein ascitic levels (<15 g/L) with advanced liver failure (Child-Pugh score > or = 9 points with serum bilirubin level > or = 3 mg/dL) or impaired renal function (serum creatinine level > or = 1.2 mg/dL, blood urea nitrogen level > or = 25 mg/dL, or serum sodium level < or = 130 mEq/L) were included in a randomized controlled trial aimed at comparing norfloxacin (35 patients) vs placebo (33 patients) in the primary prophylaxis of spontaneous bacterial peritonitis. The main end points of the trial were 3-month and 1-year probability of survival. Secondary end points were 1-year probability of development of spontaneous bacterial peritonitis and hepatorenal syndrome. **RESULTS:** Norfloxacin administration reduced the 1-year probability of developing spontaneous bacterial peritonitis (7% vs 61%, $P < .001$) and hepatorenal syndrome (28% vs 41%, $P = .02$), and improved the 3-month (94% vs 62%, $P = .003$) and the 1-year (60% vs 48%, $P = .05$) probability of survival compared with placebo. **CONCLUSIONS:** Primary prophylaxis with norfloxacin has a great impact in the clinical course of patients with advanced cirrhosis. It reduces the incidence of spontaneous bacterial peritonitis, delays the development of hepatorenal syndrome, and improves survival.

Question: Therapy

Study Type: RCT

Learner Level: Beginner

Notes: Good to teach NNT and ARR

Franklin, D., et al. (2018). "A Randomized Trial of High-Flow Oxygen Therapy in Infants with Bronchiolitis." *N Engl J Med* 378(12): 1121-1131.

BACKGROUND: High-flow oxygen therapy through a nasal cannula has been increasingly used in infants with bronchiolitis, despite limited high-quality evidence of its efficacy. The efficacy of high-flow oxygen therapy through a nasal cannula in settings other than intensive care units (ICUs) is unclear. **METHODS:** In this multicenter, randomized, controlled trial, we assigned infants younger than 12 months of age who had bronchiolitis and a need for supplemental oxygen therapy to receive either high-flow oxygen therapy (high-flow group) or standard oxygen therapy (standard-therapy group). Infants in the standard-therapy group could receive rescue high-flow oxygen therapy if their condition met criteria for treatment failure. The primary outcome was escalation of care due to treatment failure (defined as meeting ≥ 3 of 4 clinical criteria: persistent tachycardia, tachypnea, hypoxemia, and medical review triggered by a hospital early-warning tool). Secondary outcomes included duration of hospital stay, duration of oxygen therapy, and rates of transfer to a tertiary hospital, ICU admission, intubation, and adverse events. **RESULTS:** The analyses included 1472 patients. The percentage of infants receiving escalation of care was 12% (87 of 739 infants) in the high-flow group, as compared with 23% (167 of 733) in the standard-therapy group (risk difference, -11 percentage points; 95% confidence interval, -15 to -7; $P < 0.001$). No significant differences were observed in the duration of hospital stay or the duration of oxygen therapy. In each group, one case of pneumothorax ($< 1\%$ of infants) occurred. Among the 167 infants in the standard-therapy group who had treatment failure, 102 (61%) had a response to high-flow rescue therapy. **CONCLUSIONS:** Among infants with bronchiolitis who were treated outside an ICU, those who received high-flow oxygen therapy had significantly lower rates of escalation of care due to treatment failure than those in the group that received standard oxygen therapy. (Funded by the National Health and Medical Research Council and others; Australian and New Zealand Clinical Trials Registry number, ACTRN12613000388718 .).

Question: Therapy

Study Type: RCT

Learner Level: Intermediate

Notes: A good overview of what makes a good RCT, use of Kaplan-Meier curves, great comparison of the implications of using risk ratios vs risk differences.

Gerstein, H. C., et al. (2008). "Effects of intensive glucose lowering in type 2 diabetes (ACCORD Study)." *N Engl J Med* 358(24): 2545-2559.

BACKGROUND: Epidemiologic studies have shown a relationship between glycated hemoglobin levels and cardiovascular events in patients with type 2 diabetes. We investigated whether intensive therapy to target normal glycated hemoglobin levels would reduce cardiovascular events in patients with type 2 diabetes who had either established cardiovascular disease or additional cardiovascular risk factors. **METHODS:** In this randomized study, 10,251 patients (mean age, 62.2 years) with a median glycated hemoglobin level of 8.1% were assigned to receive intensive therapy (targeting a glycated hemoglobin level below 6.0%) or standard therapy (targeting a level from 7.0 to 7.9%). Of these patients, 38% were women, and 35% had had a previous cardiovascular event. The primary outcome was a composite of nonfatal myocardial infarction, nonfatal stroke, or death from cardiovascular causes. The finding of higher mortality in the intensive-therapy group led to a discontinuation of intensive therapy after a mean of 3.5 years of follow-up. **RESULTS:** At 1 year, stable median glycated hemoglobin levels of 6.4% and 7.5% were achieved in the intensive-therapy group and the standard-therapy group, respectively. During follow-up, the primary outcome occurred in 352 patients in the intensive-therapy group, as compared with 371 in the standard-therapy group (hazard ratio, 0.90; 95% confidence interval [CI], 0.78 to 1.04; $P = 0.16$). At the same time, 257 patients in the intensive-therapy group died, as compared with 203 patients in the standard-therapy group (hazard ratio, 1.22; 95% CI, 1.01 to 1.46; $P = 0.04$). Hypoglycemia requiring assistance and weight gain of more than 10 kg were more frequent in the intensive-therapy group ($P < 0.001$). **CONCLUSIONS:** As compared with standard therapy, the use of intensive therapy to target normal glycated hemoglobin levels for 3.5 years increased mortality and did not significantly reduce major cardiovascular events. These findings identify a previously unrecognized harm of intensive glucose lowering in high-risk patients with type 2 diabetes. (ClinicalTrials.gov number, NCT00000620.)

Question: Therapy

Study Type: RCT

Learner Level: Intermediate

Notes: Clearly reported methods that allow discussion of RCT validity criteria. This is one of several papers to come out within a 6 month period that does not support more aggressive glucose management. This study was stopped early for concerns about harm in the intensive insulin therapy group. You can discuss subgroup analysis (figure 3) (see also Duckworth et al. *NEJM* 2009; 360:129-39.)

Giuliano, A. E., et al. (2011). "Axillary dissection vs no axillary dissection in women with invasive breast cancer and sentinel node metastasis: a randomized clinical trial." *JAMA* 305(6): 569-575.

CONTEXT: Sentinel lymph node dissection (SLND) accurately identifies nodal metastasis of early breast cancer, but it is not clear whether further nodal dissection affects survival. OBJECTIVE: To determine the effects of complete axillary lymph node dissection (ALND) on survival of patients with sentinel lymph node (SLN) metastasis of breast cancer. DESIGN, SETTING, AND PATIENTS: The American College of Surgeons Oncology Group Z0011 trial, a phase 3 noninferiority trial conducted at 115 sites and enrolling patients from May 1999 to December 2004. Patients were women with clinical T1-T2 invasive breast cancer, no palpable adenopathy, and 1 to 2 SLNs containing metastases identified by frozen section, touch preparation, or hematoxylin-eosin staining on permanent section. Targeted enrollment was 1900 women with final analysis after 500 deaths, but the trial closed early because mortality rate was lower than expected. INTERVENTIONS: All patients underwent lumpectomy and tangential whole-breast irradiation. Those with SLN metastases identified by SLND were randomized to undergo ALND or no further axillary treatment. Those randomized to ALND underwent dissection of 10 or more nodes. Systemic therapy was at the discretion of the treating physician. MAIN OUTCOME MEASURES: Overall survival was the primary end point, with a noninferiority margin of a 1-sided hazard ratio of less than 1.3 indicating that SLND alone is noninferior to ALND. Disease-free survival was a secondary end point. RESULTS: Clinical and tumor characteristics were similar between 445 patients randomized to ALND and 446 randomized to SLND alone. However, the median number of nodes removed was 17 with ALND and 2 with SLND alone. At a median follow-up of 6.3 years (last follow-up, March 4, 2010), 5-year overall survival was 91.8% (95% confidence interval [CI], 89.1%-94.5%) with ALND and 92.5% (95% CI, 90.0%-95.1%) with SLND alone; 5-year disease-free survival was 82.2% (95% CI, 78.3%-86.3%) with ALND and 83.9% (95% CI, 80.2%-87.9%) with SLND alone. The hazard ratio for treatment-related overall survival was 0.79 (90% CI, 0.56-1.11) without adjustment and 0.87 (90% CI, 0.62-1.23) after adjusting for age and adjuvant therapy. CONCLUSION: Among patients with limited SLN metastatic breast cancer treated with breast conservation and systemic therapy, the use of SLND alone compared with ALND did not result in inferior survival. TRIAL REGISTRATION: clinicaltrials.gov Identifier: NCT00003855.

Question: Therapy

Study Type: RCT

Learner Level: Intermediate - Advanced

Notes: This is a well done RCT overall, could be useful as a therapy article, but its strength as a teaching article is in the non-inferiority design, and the pitfalls of non-inferiority when you don't see the number of outcomes you planned for. Intermediate to advanced

Granger, C. B., et al. (2011). "Apixaban versus warfarin in patients with atrial fibrillation." *N Engl J Med* 365(11): 981-992.

BACKGROUND: Vitamin K antagonists are highly effective in preventing stroke in patients with atrial fibrillation but have several limitations. Apixaban is a novel oral direct factor Xa inhibitor that has been shown to reduce the risk of stroke in a similar population in comparison with aspirin. METHODS: In this randomized, double-blind trial, we compared apixaban (at a dose of 5 mg twice daily) with warfarin (target international normalized ratio, 2.0 to 3.0) in 18,201 patients with atrial fibrillation and at least one additional risk factor for stroke. The primary outcome was ischemic or hemorrhagic stroke or systemic embolism. The trial was designed to test for noninferiority, with key secondary objectives of testing for superiority with respect to the primary outcome and to the rates of major bleeding and death from any cause. RESULTS: The median duration of follow-up was 1.8 years. The rate of the primary outcome was 1.27% per year in the apixaban group, as compared with 1.60% per year in the warfarin group (hazard ratio with apixaban, 0.79; 95% confidence interval [CI], 0.66 to 0.95; $P < 0.001$ for noninferiority; $P = 0.01$ for superiority). The rate of major bleeding was 2.13% per year in the apixaban group, as compared with 3.09% per year in the warfarin group (hazard ratio, 0.69; 95% CI, 0.60 to 0.80; $P < 0.001$), and the rates of death from any cause were 3.52% and 3.94%, respectively (hazard ratio, 0.89; 95% CI, 0.80 to 0.99; $P = 0.047$). The rate of hemorrhagic stroke was 0.24% per year in the apixaban group, as compared with 0.47% per year in the warfarin group (hazard ratio, 0.51; 95% CI, 0.35 to 0.75; $P < 0.001$), and the rate of ischemic or uncertain type of stroke was 0.97% per year in the apixaban group and 1.05% per year in the warfarin group (hazard ratio, 0.92; 95% CI, 0.74 to 1.13; $P = 0.42$). CONCLUSIONS: In patients with atrial fibrillation, apixaban was superior to warfarin in preventing stroke or systemic embolism, caused less bleeding, and resulted in lower mortality. (Funded by Bristol-Myers Squibb and Pfizer; ARISTOTLE ClinicalTrials.gov number, NCT00412984.)

Question: Therapy

Study Type: RCT

Learner Level: Advanced

Notes: noninferiority with Apixaban for Afib - outstanding non-inferiority paper to review NI margins, methodology, assumptions

Hahn, D. L. (2011). "Azithromycin for prevention of exacerbations of COPD." *N Engl J Med* 365(23): 2236-2236; author reply 2236-2237.

Question: Therapy

Study Type: RCT

Learner Level: Beginner

Notes: a good article to demonstrate basics of assessing validity of a therapeutic trial. Easy.

Halkes, P. H., et al. (2006). "Aspirin plus dipyridamole versus aspirin alone after cerebral ischaemia of arterial origin (ESPRIT): randomised controlled trial." *Lancet* 367(9523): 1665-1673.

BACKGROUND: Results of trials of aspirin and dipyridamole combined versus aspirin alone for the secondary prevention of vascular events after ischaemic stroke of presumed arterial origin are inconsistent. Our aim was to resolve this uncertainty. **METHODS:** We did a randomised controlled trial in which we assigned patients to aspirin (30-325 mg daily) with (n=1363) or without (n=1376) dipyridamole (200 mg twice daily) within 6 months of a transient ischaemic attack or minor stroke of presumed arterial origin. Our primary outcome event was the composite of death from all vascular causes, non-fatal stroke, non-fatal myocardial infarction, or major bleeding complication, whichever happened first. Treatment was open, but auditing of outcome events was blinded. Primary analysis was by intention to treat. This study is registered as an International Standard Randomised Controlled Trial (number ISRCTN73824458) and with (NCT00161070). **FINDINGS:** Mean follow-up was 3.5 years (SD 2.0). Median aspirin dose was 75 mg in both treatment groups (range 30-325); extended-release dipyridamole was used by 83% (n=1131) of patients on the combination regimen. Primary outcome events arose in 173 (13%) patients on aspirin and dipyridamole and in 216 (16%) on aspirin alone (hazard ratio 0.80, 95% CI 0.66-0.98; absolute risk reduction 1.0% per year, 95% CI 0.1-1.8). Addition of the ESPRIT data to the meta-analysis of previous trials resulted in an overall risk ratio for the composite of vascular death, stroke, or myocardial infarction of 0.82 (95% CI 0.74-0.91). Patients on aspirin and dipyridamole discontinued trial medication more often than those on aspirin alone (470 vs 184), mainly because of headache. **INTERPRETATION:** The ESPRIT results, combined with the results of previous trials, provide sufficient evidence to prefer the combination regimen of aspirin plus dipyridamole over aspirin alone as antithrombotic therapy after cerebral ischaemia of arterial origin.

Question: Therapy

Study Type: RCT

Learner Level: Beginner

Notes: Good to teach composite outcomes

Hebert, P. C., et al. (1999). "A multicenter, randomized, controlled clinical trial of transfusion requirements in critical care. Transfusion Requirements in Critical Care Investigators, Canadian Critical Care Trials Group." *N Engl J Med* 340(6): 409-417.

BACKGROUND: To determine whether a restrictive strategy of red-cell transfusion and a liberal strategy produced equivalent results in critically ill patients, we compared the rates of death from all causes at 30 days and the severity of organ dysfunction. **METHODS:** We enrolled 838 critically ill patients with euvoemia after initial treatment who had hemoglobin concentrations of less than 9.0 g per deciliter within 72 hours after admission to the intensive care unit and randomly assigned 418 patients to a restrictive strategy of transfusion, in which red cells were transfused if the hemoglobin concentration dropped below 7.0 g per deciliter and hemoglobin concentrations were maintained at 7.0 to 9.0 g per deciliter, and 420 patients to a liberal strategy, in which transfusions were given when the hemoglobin concentration fell below 10.0 g per deciliter and hemoglobin concentrations were maintained at 10.0 to 12.0 g per deciliter. **RESULTS:** Overall, 30-day mortality was similar in the two groups (18.7 percent vs. 23.3 percent, $P=0.11$). However, the rates were significantly lower with the restrictive transfusion strategy among patients who were less acutely ill -- those with an Acute Physiology and Chronic Health Evaluation II score of $< \text{or} = 20$ (8.7 percent in the restrictive-strategy group and 16.1 percent in the liberal-strategy group; $P=0.03$) -- and among patients who were less than 55 years of age (5.7 percent and 13.0 percent, respectively; $P=0.02$), but not among patients with clinically significant cardiac disease (20.5 percent and 22.9 percent, respectively; $P=0.69$). The mortality rate during hospitalization was significantly lower in the restrictive-strategy group (22.3 percent vs. 28.1 percent, $P=0.05$). **CONCLUSIONS:** A restrictive strategy of red-cell transfusion is at least as effective as and possibly superior to a liberal transfusion strategy in critically ill patients, with the possible exception of patients with acute myocardial infarction and unstable angina.

Question: Therapy

Study Type: RCT

Learner Level: Intermediate

Notes: An RCT of restrictive RBC transfusion strategy vs. liberal RBC transfusion strategy. Clear methods well reported. Good paper to discuss the concept of an equivalency trial. While you cannot calculate an NNT for the main results (as they show no difference), you can calculate an NNT for in hospital mortality which was different between groups. Good paper to combine with Rao et al JAMA 2004; 292:155-1562

Hochman, J. S., et al. (2006). "Coronary intervention for persistent occlusion after myocardial infarction." *N Engl J Med* 355(23): 2395-2407.

BACKGROUND: It is unclear whether stable, high-risk patients with persistent total occlusion of the infarct-related coronary artery identified after the currently accepted period for myocardial salvage has passed should undergo percutaneous coronary intervention (PCI) in addition to receiving optimal medical therapy to reduce the risk of subsequent events. **METHODS:** We conducted a randomized study involving 2166 stable patients who had total occlusion of the infarct-related artery 3 to 28 days after myocardial infarction and who met a high-risk criterion (an ejection fraction of $< 50\%$ or proximal occlusion). Of these patients, 1082 were assigned to routine PCI and stenting with optimal medical therapy, and 1084 were assigned to optimal medical therapy alone. The primary end point was a composite of death, myocardial reinfarction, or New York Heart Association (NYHA) class IV heart failure. **RESULTS:** The 4-year cumulative primary event rate was 17.2% in the PCI group and 15.6% in the medical therapy group (hazard ratio for death, reinfarction, or heart failure in the PCI group as compared with the medical therapy group, 1.16; 95% confidence interval

[CI], 0.92 to 1.45; P=0.20). Rates of myocardial reinfarction (fatal and nonfatal) were 7.0% and 5.3% in the two groups, respectively (hazard ratio, 1.36; 95% CI, 0.92 to 2.00; P=0.13). Rates of nonfatal reinfarction were 6.9% and 5.0%, respectively (hazard ratio, 1.44; 95% CI, 0.96 to 2.16; P=0.08); only six reinfarctions (0.6%) were related to assigned PCI procedures. Rates of NYHA class IV heart failure (4.4% vs. 4.5%) and death (9.1% vs. 9.4%) were similar. There was no interaction between treatment effect and any subgroup variable (age, sex, race or ethnic group, infarct-related artery, ejection fraction, diabetes, Killip class, and the time from myocardial infarction to randomization). CONCLUSIONS: PCI did not reduce the occurrence of death, reinfarction, or heart failure, and there was a trend toward excess reinfarction during 4 years of follow-up in stable patients with occlusion of the infarct-related artery 3 to 28 days after myocardial infarction. (ClinicalTrials.gov number, NCT00004562 [ClinicalTrials.gov]).

Question: Therapy

Study Type: RCT

Learner Level: .

Notes: Great teaching points

Hollingsworth, J. M., et al. (2006). "Medical therapy to facilitate urinary stone passage: a meta-analysis." *Lancet* 368(9542): 1171-1179.

BACKGROUND: Medical therapies to ease urinary-stone passage have been reported, but are not generally used. If effective, such therapies would increase the options for treatment of urinary stones. To assess efficacy, we sought to identify and summarise all randomised controlled trials in which calcium-channel blockers or alpha blockers were used to treat urinary stone disease. **METHODS:** We searched MEDLINE, Pre-MEDLINE, CINAHL, and EMBASE, as well as scientific meeting abstracts, up to July, 2005. All randomised controlled trials in which calcium-channel blockers or alpha blockers were used to treat ureteral stones were eligible for inclusion in our analysis. Data from nine trials (number of patients=693) were pooled. The main outcome was the proportion of patients who passed stones. We calculated the summary estimate of effect associated with medical therapy use using random-effects and fixed-effects models. **FINDINGS:** Patients given calcium-channel blockers or alpha blockers had a 65% (absolute risk reduction=0.31 95% CI 0.25-0.38) greater likelihood of stone passage than those not given such treatment (pooled risk ratio 1.65; 95% CI 1.45-1.88). The pooled risk ratio for alpha blockers was 1.54 (1.29-1.85) and for calcium-channel blockers with steroids was 1.90 (1.51-2.40). The proportion of heterogeneity not explained by chance alone was 28%. The number needed to treat was 4. **INTERPRETATION:** Although a high-quality randomised trial is necessary to confirm its efficacy, our findings suggest that medical therapy is an option for facilitation of urinary-stone passage for patients amenable to conservative management, potentially obviating the need for surgery.

Question: Therapy

Study Type: Meta-analysis

Learner Level: Beginner

Notes: The study pools relatively few studies of low methodological quality. Suitable for group exercises in which learners abstract data assess methodological quality of individual RCTs and then pool data.

Hooton, T. M., et al. (2018). "Effect of Increased Daily Water Intake in Premenopausal Women With Recurrent Urinary Tract Infections: A Randomized Clinical Trial." *JAMA Intern Med* 178(11): 1509-1515.

Importance: Increased hydration is often recommended as a preventive measure for women with recurrent cystitis, but supportive data are sparse. **Objective:** To assess the efficacy of increased daily water intake on the frequency of recurrent cystitis in premenopausal women. **Design, Setting, and Participants:** Randomized, open-label, controlled, 12-month trial at a clinical research center (years 2013-2016). Among 163 healthy women with recurrent cystitis (>=3 episodes in past year) drinking less than 1.5 L of fluid daily assessed for eligibility, 23 were excluded and 140 assigned to water or control group. Assessments of daily fluid intake, urinary hydration, and cystitis symptoms were performed at baseline, 6- and 12-month visits, and monthly telephone calls. **Interventions:** Participants were randomly assigned to drink, in addition to their usual fluid intake, 1.5 L of water daily (water group) or no additional fluids (control group) for 12 months. **Main Outcomes and Measures:** Primary outcome measure was frequency of recurrent cystitis over 12 months. Secondary outcomes were number of antimicrobial regimens used, mean time interval between cystitis episodes, and 24-hour urinary hydration measurements. **Results:** The mean (SD) age of the 140 participants was 35.7 (8.4) years, and the mean (SD) number of cystitis episodes in the previous year was 3.3 (0.6). During the 12-month study period, the mean (SD) number of cystitis episodes was 1.7 (95% CI, 1.5-1.8) in the water group compared with 3.2 (95% CI, 3.0-3.4) in the control group, with a difference in means of 1.5 (95% CI, 1.2-1.8; P < .001). Overall, there were 327 cystitis episodes, 111 in the water group and 216 in the control group. The mean number of antimicrobial regimens used to treat cystitis episodes was 1.9 (95% CI, 1.7-2.2) and 3.6 (95% CI, 3.3-4.0), respectively, with a difference in means of 1.7 (95% CI, 1.3-2.1; P < .001). The mean time interval between cystitis episodes was 142.8 (95% CI, 127.4-160.1) and 84.4 (95% CI, 75.4-94.5) days, respectively, with a difference in means of 58.4 (95% CI, 39.4-77.4; P < .001). Between baseline and 12 months, participants in the water group, compared with those in the control group, had increased mean (SD) urine volume (1.4 [0.04] vs 0.1 [0.04] L; P < .001) and voids (2.4 [0.2] vs -0.1 [0.2]; P < .001) and decreased urine osmolality (-402.8 [19.6] vs -24.0 [19.5] mOsm/kg; P < .001). **Conclusions and Relevance:** Increased water intake is an effective antimicrobial-sparing strategy to prevent recurrent cystitis in premenopausal women at high risk for recurrence who drink low volumes of fluid daily. **Trial Registration:** ClinicalTrials.gov identifier: NCT02444975.

Question: Therapy

Study Type: RCT

Learner Level: Beginner

Notes: A good starter RCT of a relevant question to many people – drinking water to decrease repeat UTI in young women. It was a well-designed trial that might be good for early learners who want to focus on the critical appraisal of an RCT. There is an ACP journal club which can also help for new teachers as they gain confidence in appraising study.

Hull, R. D., et al. (2001). "Extended out-of-hospital low-molecular-weight heparin prophylaxis against deep venous thrombosis in patients after elective hip arthroplasty: a systematic review." *Ann Intern Med* 135(10): 858-869.

PURPOSE: Evidence-based medicine guidelines based on venographic end points recommend in-hospital prophylaxis with low-molecular-weight heparin (LMWH) in patients having elective hip surgery. Emerging data suggest that out-of-hospital use may offer additional protection; however, uncertainty remains about the risk-benefit ratio. To provide clinicians with a practical pathway for translating clinical research into practice, we systematically reviewed trials comparing extended out-of-hospital LMWH prophylaxis versus placebo. **DATA SOURCES:** Studies were identified by 1) searching PubMed, MEDLINE, and the Cochrane Library Database for reports published from January 1976 to May 2001; 2) reviewing references from retrieved articles; 3) scanning abstracts from conference proceedings; and 4) contacting pharmaceutical companies and investigators of the original reports. **STUDY SELECTION:** Randomized, controlled trials comparing extended out-of-hospital prophylaxis with LMWH versus placebo in patients having elective hip arthroplasty. **DATA EXTRACTION:** Two reviewers extracted data independently. Reviewers evaluated study quality by using a validated four-item instrument. **DATA SYNTHESIS:** Six of seven original articles met the defined inclusion criteria. The included studies were double-blind trials that used proper randomization procedures. Compared with placebo, extended out-of-hospital prophylaxis decreased the frequency of all episodes of deep venous thrombosis (placebo rate, 150 of 666 patients [22.5%]; relative risk, 0.41 [95% CI, 0.32 to 0.54; $P < 0.001$]), proximal venous thrombosis (placebo rate, 76 of 678 patients [11.2%]; relative risk, 0.31 [CI, 0.20 to 0.47; $P < 0.001$]), and symptomatic venous thromboembolism (placebo rate, 36 of 862 patients [4.2%]; relative risk, 0.36 [CI, 0.20 to 0.67; $P = 0.001$]). Major bleeding was rare, occurring in only one patient in the placebo group. **CONCLUSIONS:** Extended LMWH prophylaxis showed consistent effectiveness and safety in the trials (regardless of study variations in clinical practice and length of hospital stay) for venographic deep venous thrombosis and symptomatic venous thromboembolism. The aggregate findings support the need for extended out-of-hospital prophylaxis in patients undergoing hip arthroplasty surgery.

Question: Therapy

Study Type: Meta-analysis

Learner Level: Intermediate

Notes: Important question which prompts good discussion of risks v. benefits v. cost. Good Discussion Points: Strengths: Well designed with clear description of methods. Can use methods to prompt vocabulary discussion for concepts pertaining to Meta-analysis (e.g. methods for validity assessment, summary treatment effects, sensitivity analysis, heterogeneity); Weaknesses: Variety of interventions can become very confusing to learners. More Advanced Discussion: might consider getting the QUOROM statement that is referenced in the methods of the paper (p. 859) for more in depth discussion of Meta-analysis methodology. Check list on p 1897 can be used to assess this (or any other systematic review); For more advanced discussion see also Moher 1999.

Humphrey, L. L., et al. (2002). "Postmenopausal hormone replacement therapy and the primary prevention of cardiovascular disease." *Ann Intern Med* 137(4): 273-284.

PURPOSE: To evaluate the value of hormone replacement therapy (HRT) in the primary prevention of cardiovascular disease (CVD) and coronary artery disease (CAD). **DATA SOURCES:** MEDLINE and Cochrane databases were searched for all primary prevention studies reporting CVD or CAD incidence, mortality, or both in association with HRT; reference lists, letters, editorials, and reviews were also reviewed. **DATA EXTRACTION:** All studies were reviewed, abstracted, and rated for quality. **STUDY SELECTION:** Only studies of good or fair quality, according to U.S. Preventive Services Task Force (USPSTF) criteria, were included in the detailed review and meta-analysis. **DATA SYNTHESIS:** The summary relative risk with any HRT use was 0.75 (95% credible interval [CrI], 0.42 to 1.23) for CVD mortality and 0.74 (CrI, 0.36 to 1.45) for CAD mortality. The summary relative risk with any use was 1.28 (CrI, 0.86 to 2.00) for CVD incidence and 0.87 (CrI, 0.62 to 1.21) for CAD incidence. Further analysis of studies adjusting for socioeconomic status, as well as other major CAD risk factors, showed a summary relative risk of 1.07 (CrI, 0.79 to 1.48) for CAD incidence associated with any HRT use. Similar results were found when the analysis was stratified by studies adjusting for alcohol consumption, exercise, or both, in addition to other major risk factors, suggesting confounding by these factors. **CONCLUSIONS:** This meta-analysis differs from previous meta-analyses by evaluating potential explanatory variables of the relationship between HRT, CVD, and CAD. The adjusted meta-analysis is consistent with recent randomized trials that have shown no benefit in the secondary or primary prevention of CVD events. A valid answer to the role of HRT in the primary prevention of CVD will best come from randomized, controlled trials.

Question: Therapy

Study Type: Meta-analysis

Learner Level: Intermediate

Notes: Pros: Can discuss Cochrane controlled trials register. Validity criteria for inclusion of studies into the SR. Search strategies for SRs. Cons: Long, Dated

Huss, A., et al. (2009). "Efficacy of pneumococcal vaccination in adults: a meta-analysis." *CMAJ* 180(1): 48-58.

BACKGROUND: Clinical trials and meta-analyses have produced conflicting results of the efficacy of unconjugated pneumococcal polysaccharide vaccine in adults. We sought to evaluate the vaccine's efficacy on clinical outcomes as well as the methodologic quality of the trials. **METHODS:** We searched several databases and all bibliographies of reviews and meta-analyses for clinical trials that compared pneumococcal polysaccharide vaccine with a control. We examined rates of pneumonia and death, taking the methodologic quality of the trials into consideration. **RESULTS:** We included 22 trials involving 101 507 participants: 11 trials reported on presumptive pneumococcal pneumonia, 19 on all-cause pneumonia and 12 on all-cause mortality. The current 23-valent vaccine was used in 8 trials. The relative risk (RR) was 0.64 (95% confidence interval [CI] 0.43-0.96) for presumptive pneumococcal pneumonia and 0.73 (95% CI 0.56-0.94) for all-cause pneumonia. There was significant heterogeneity between the trials reporting on presumptive pneumonia ($I^2 = 74\%$, $p < 0.001$) and between those reporting on all-cause pneumonia ($I^2 = 90\%$, $p < 0.001$). The RR for all-cause mortality was 0.97 (95% CI 0.87-1.09), with moderate heterogeneity between trials ($I^2 = 44\%$, $p = 0.053$). Trial quality, especially regarding double blinding, explained a substantial proportion of the heterogeneity in the trials reporting on presumptive pneumonia and all-cause pneumonia. There was little evidence of vaccine protection in trials of higher methodologic quality (RR 1.20, 95% CI 0.75-1.92, for presumptive pneumonia; and 1.19, 95% CI 0.95-1.49, for all-cause pneumonia in double-blind trials; p for heterogeneity > 0.05). The results for all-cause mortality in double-blind trials were similar to those in all trials combined. There was little evidence of vaccine protection among elderly patients or adults with chronic illness in analyses of all trials (RR 1.04, 95% CI 0.78-1.38, for presumptive pneumococcal pneumonia; 0.89, 95% CI 0.69-1.14, for all-cause pneumonia; and 1.00, 95% CI 0.87-1.14, for all-cause mortality). **INTERPRETATION:** Pneumococcal vaccination does not appear to be effective in preventing pneumonia, even in populations for whom the vaccine is currently recommended.

Question: Therapy

Study Type: Meta-analysis

Learner Level: Beginner

Notes: article is suited to describe the methodology, advantages and limitations of a meta-analysis. Also useful for the purposes of recognizing the impact of primary trial heterogeneity on the results of a meta-analysis.

Jack, B. W., et al. (2009). "A reengineered hospital discharge program to decrease rehospitalization: a randomized trial." *Ann Intern Med* 150(3): 178-187.

BACKGROUND: Emergency department visits and rehospitalization are common after hospital discharge. **OBJECTIVE:** To test the effects of an intervention designed to minimize hospital utilization after discharge. **DESIGN:** Randomized trial using block randomization of 6 and 8. Randomly arranged index cards were placed in opaque envelopes labeled consecutively with study numbers, and participants were assigned a study group by revealing the index card. **SETTING:** General medical service at an urban, academic, safety-net hospital. **PATIENTS:** 749 English-speaking hospitalized adults (mean age, 49.9 years). **INTERVENTION:** A nurse discharge advocate worked with patients during their hospital stay to arrange follow-up appointments, confirm medication reconciliation, and conduct patient education with an individualized instruction booklet that was sent to their primary care provider. A clinical pharmacist called patients 2 to 4 days after discharge to reinforce the discharge plan and review medications. Participants and providers were not blinded to treatment assignment. **MEASUREMENTS:** Primary outcomes were emergency department visits and hospitalizations within 30 days of discharge. Secondary outcomes were self-reported preparedness for discharge and frequency of primary care providers' follow-up within 30 days of discharge. Research staff doing follow-up were blinded to study group assignment. **RESULTS:** Participants in the intervention group ($n = 370$) had a lower rate of hospital utilization than those receiving usual care ($n = 368$) (0.314 vs. 0.451 visit per person per month; incidence rate ratio, 0.695 [95% CI, 0.515 to 0.937]; $P = 0.009$). The intervention was most effective among participants with hospital utilization in the 6 months before index admission ($P = 0.014$). Adverse events were not assessed; these data were collected but are still being analyzed. **LIMITATION:** This was a single-center study in which not all potentially eligible patients could be enrolled, and outcome assessment sometimes relied on participant report. **CONCLUSION:** A package of discharge services reduced hospital utilization within 30 days of discharge. **FUNDING:** Agency for Healthcare Research and Quality and National Heart, Lung, and Blood Institute, National Institutes of Health.

Question: Therapy

Study Type: RCT

Learner Level: Beginner

Notes: Great for teaching principles of RCT methodology. Specifically good for teaching 1) randomization : sequence generation (blocked randomization scheme) and allocation concealment (opaque envelopes as well as varied block sizes); 2) blinding (who can be blinded when the intervention is a process enhancement)

James, P. A., et al. (2014). "2014 evidence-based guideline for the management of high blood pressure in adults: report from the panel members appointed to the Eighth Joint National Committee (JNC 8)." *JAMA* 311(5): 507-520.

Hypertension is the most common condition seen in primary care and leads to myocardial infarction, stroke, renal failure, and death if not detected early and treated appropriately. Patients want to be assured that blood pressure (BP) treatment will reduce their disease burden, while clinicians want guidance on hypertension management using the best scientific evidence. This report takes a rigorous, evidence-based approach to recommend treatment thresholds, goals, and medications in the management of hypertension in adults. Evidence was drawn from randomized controlled trials, which represent the gold standard for determining

efficacy and effectiveness. Evidence quality and recommendations were graded based on their effect on important outcomes. There is strong evidence to support treating hypertensive persons aged 60 years or older to a BP goal of less than 150/90 mm Hg and hypertensive persons 30 through 59 years of age to a diastolic goal of less than 90 mm Hg; however, there is insufficient evidence in hypertensive persons younger than 60 years for a systolic goal, or in those younger than 30 years for a diastolic goal, so the panel recommends a BP of less than 140/90 mm Hg for those groups based on expert opinion. The same thresholds and goals are recommended for hypertensive adults with diabetes or nondiabetic chronic kidney disease (CKD) as for the general hypertensive population younger than 60 years. There is moderate evidence to support initiating drug treatment with an angiotensin-converting enzyme inhibitor, angiotensin receptor blocker, calcium channel blocker, or thiazide-type diuretic in the nonblack hypertensive population, including those with diabetes. In the black hypertensive population, including those with diabetes, a calcium channel blocker or thiazide-type diuretic is recommended as initial therapy. There is moderate evidence to support initial or add-on antihypertensive therapy with an angiotensin-converting enzyme inhibitor or angiotensin receptor blocker in persons with CKD to improve kidney outcomes. Although this guideline provides evidence-based recommendations for the management of high BP and should meet the clinical needs of most patients, these recommendations are not a substitute for clinical judgment, and decisions about care must carefully consider and incorporate the clinical characteristics and circumstances of each individual patient.

Question: Therapy

Study Type: Guideline

Learner Level: Beginner

Notes: Type: Clinical Practice Guideline. Teaching points: important, well done, evidence based guideline, transparently reported. The results are controversial as there was not agreement on the panel for recommendations for BP cut offs in individuals 60 or older (without DM or renal disease). The nice part about this study is that it makes very clear links to RCT-level evidence and limits its discussion and recommendations to that which can be clearly linked to high quality evidence. Learners can discuss the pluses and minuses of that approach. As an extra added benefit: it is very hard to find a CPG that is concise enough to print for teaching and this one is.

Johnson, D. W., et al. (1998). "A comparison of nebulized budesonide, intramuscular dexamethasone, and placebo for moderately severe croup." *N Engl J Med* 339(8): 498-503.

BACKGROUND: In children with croup, treatment with nebulized budesonide decreases symptoms, but it is uncertain how budesonide compares with dexamethasone, the conventional therapy for croup, and whether either reduces the rate of hospitalization. **METHODS:** We performed a double-blind, randomized trial involving 144 children with moderately severe croup. The children were treated with racemic epinephrine and a single dose of 4 mg of nebulized budesonide (48 children), 0.6 mg of intramuscular dexamethasone per kilogram of body weight (47 children), or placebo (49 children). The children were assessed before treatment and then hourly for five hours after treatment. Physicians who were unaware of the treatment assignments determined the children's need for further treatment and hospitalization. **RESULTS:** The characteristics of the groups were similar at base line, including the types of viruses identified, the types of croup, and the clinical severity of the illness. The overall rates of hospitalization were 71 percent in the placebo group (35 of 49 children), 38 percent in the budesonide group (18 of 48 children), and 23 percent in the dexamethasone group (11 of 47 children) (unadjusted $P=0.001$ for the comparison of budesonide with placebo, $P<0.001$ for the comparison of dexamethasone with placebo, and $P=0.18$ for the comparison of budesonide with dexamethasone). Children treated with budesonide or dexamethasone had a greater improvement in croup scores than those given placebo ($P=0.03$ and $P<0.001$, respectively), and those treated with dexamethasone had a greater improvement than those treated with budesonide ($P=0.003$). **CONCLUSIONS:** In children with moderately severe croup, treatment with intramuscular dexamethasone or nebulized budesonide resulted in more rapid clinical improvement than did the administration of placebo, with dexamethasone offering the greatest improvement. Treatment with either glucocorticoid resulted in fewer hospitalizations.

Question: Therapy

Study Type: RCT

Learner Level: Beginner / Intermediate

Notes: Fabulous methodology, clearly written; Easy NNT calculations. Note: this paper is of relevance to pediatricians, but also to anyone who has ever had a child with croup! Good discussion points: Awesome paper for discussion of blinding and the challenges that come up; Intention to treat analysis; For those who might want a greater challenge, you can discuss sample size calculations and type I error (page 500); Could address some issues of ethics including study termination

Johnston, S. C., et al. (2018). "Clopidogrel and Aspirin in Acute Ischemic Stroke and High-Risk TIA." *N Engl J Med* 379(3): 215-225.

BACKGROUND: Combination antiplatelet therapy with clopidogrel and aspirin may reduce the rate of recurrent stroke during the first 3 months after a minor ischemic stroke or transient ischemic attack (TIA). A trial of combination antiplatelet therapy in a Chinese population has shown a reduction in the risk of recurrent stroke. We tested this combination in an international population. **METHODS:** In a randomized trial, we assigned patients with minor ischemic stroke or high-risk TIA to receive either clopidogrel at a loading dose of 600 mg on day 1, followed by 75 mg per day, plus aspirin (at a dose of 50 to 325 mg per day) or the same range of doses of aspirin alone. The dose of aspirin in each group was selected by the site investigator. The primary efficacy outcome in a time-to-event analysis was the risk of a composite of major ischemic events, which was defined as ischemic stroke, myocardial infarction, or death from an ischemic vascular event, at 90 days. **RESULTS:** A total of 4881 patients were enrolled at 269 international sites. The trial was halted after 84% of the anticipated number of patients had been enrolled because the data and safety monitoring

board had determined that the combination of clopidogrel and aspirin was associated with both a lower risk of major ischemic events and a higher risk of major hemorrhage than aspirin alone at 90 days. Major ischemic events occurred in 121 of 2432 patients (5.0%) receiving clopidogrel plus aspirin and in 160 of 2449 patients (6.5%) receiving aspirin plus placebo (hazard ratio, 0.75; 95% confidence interval [CI], 0.59 to 0.95; P=0.02), with most events occurring during the first week after the initial event. Major hemorrhage occurred in 23 patients (0.9%) receiving clopidogrel plus aspirin and in 10 patients (0.4%) receiving aspirin plus placebo (hazard ratio, 2.32; 95% CI, 1.10 to 4.87; P=0.02). CONCLUSIONS: In patients with minor ischemic stroke or high-risk TIA, those who received a combination of clopidogrel and aspirin had a lower risk of major ischemic events but a higher risk of major hemorrhage at 90 days than those who received aspirin alone. (Funded by the National Institute of Neurological Disorders and Stroke; POINT ClinicalTrials.gov number, NCT00991029 .).

Question: Therapy

Study Type: RCT

Learner Level: Beginner

Notes: Good article to discuss clinical relevance and reinforce periods of highest risk.

Juni, P., et al. (2004). "Risk of cardiovascular events and rofecoxib: cumulative meta-analysis." *Lancet* 364(9450): 2021-2029.

BACKGROUND: The cyclo-oxygenase 2 inhibitor rofecoxib was recently withdrawn because of cardiovascular adverse effects. An increased risk of myocardial infarction had been observed in 2000 in the Vioxx Gastrointestinal Outcomes Research study (VIGOR), but was attributed to cardioprotection of naproxen rather than a cardiotoxic effect of rofecoxib. We used standard and cumulative random-effects meta-analyses of randomised controlled trials and observational studies to establish whether robust evidence on the adverse effects of rofecoxib was available before September, 2004. **METHODS:** We searched bibliographic databases and relevant files of the US Food and Drug Administration. We included all randomised controlled trials in patients with chronic musculoskeletal disorders that compared rofecoxib with other non-steroidal anti-inflammatory drugs (NSAIDs) or placebo, and cohort and case-control studies of cardiovascular risk and naproxen. Myocardial infarction was the primary endpoint. **FINDINGS:** We identified 18 randomised controlled trials and 11 observational studies. By the end of 2000 (52 myocardial infarctions, 20742 patients) the relative risk from randomised controlled trials was 2.30 (95% CI 1.22-4.33, p=0.010), and 1 year later (64 events, 21432 patients) it was 2.24 (1.24-4.02, p=0.007). There was little evidence that the relative risk differed depending on the control group (placebo, non-naproxen NSAID, or naproxen; p=0.41) or trial duration (p=0.82). In observational studies, the cardioprotective effect of naproxen was small (combined estimate 0.86 [95% CI 0.75-0.99]) and could not have explained the findings of the VIGOR trial. **INTERPRETATION:** Our findings indicate that rofecoxib should have been withdrawn several years earlier. The reasons why manufacturer and drug licensing authorities did not continuously monitor and summarise the accumulating evidence need to be clarified.

Question: Therapy (Harm)

Study Type: Meta-analysis

Learner Level: Intermediate

Notes: Topical question of broad clinical interest which learners find interesting. Robust cumulative meta-analysis showing the chronology of relevant RCTs on the topic demonstrating harm of this commonly prescribed medication.

Keeney, G. E., et al. (2014). "Dexamethasone for acute asthma exacerbations in children: a meta-analysis." *Pediatrics* 133(3): 493-499.

BACKGROUND AND OBJECTIVE: Dexamethasone has been proposed as an equivalent therapy to prednisone/prednisolone for acute asthma exacerbations in pediatric patients. Although multiple small trials exist, clear consensus data are lacking. This systematic review and meta-analysis aimed to determine whether intramuscular or oral dexamethasone is equivalent or superior to a 5-day course of oral prednisone or prednisolone. The primary outcome of interest was return visits or hospital readmissions. **METHODS:** A search of PubMed (Medline) through October 19, 2013, by using the keywords dexamethasone or decadron and asthma or status asthmaticus identified potential studies. Six randomized controlled trials in the emergency department of children ≤ 18 years of age comparing dexamethasone with prednisone/prednisolone for the treatment of acute asthma exacerbations were included. Data were abstracted by 4 authors and verified by a second author. Two reviewers evaluated study quality independently and interrater agreement was assessed. **RESULTS:** There was no difference in relative risk (RR) of relapse between the 2 groups at any time point (5 days RR 0.90, 95% confidence interval [CI] 0.46-1.78, Q = 1.86, df = 3, I² = 0.0%, 10-14 days RR 1.14, 95% CI 0.77-1.67, Q = 0.84, df = 2, I² = 0.0%, or 30 days RR 1.20, 95% CI 0.03-56.93). Patients who received dexamethasone were less likely to experience vomiting in either the emergency department (RR 0.29, 95% CI 0.12-0.69, Q = 3.78, df = 3, I² = 20.7%) or at home (RR 0.32, 95% CI 0.14-0.74, Q = 2.09, df = 2, I² = 4.2%). **CONCLUSIONS:** Practitioners should consider single or 2-dose regimens of dexamethasone as a viable alternative to a 5-day course of prednisone/prednisolone.

Question: Therapy

Study Type: Meta-Analysis

Learner Level: Beginner/Intermediate

Notes: Not the greatest meta-analysis overall (with regards to should these article be combined) but has good F's: frame it, find it, filter/fetch it, forrest plots. Can have good discussion regarding the ultimate analysis of the results – should these studies have been combined and analyzed by meta-analysis or simply reported as a systematic review? 6 studies – 3 used intramuscular dexamethasone and 3 used oral. Of the 3 with oral dex, 2 used 1 daily dose on 2 consecutive days (1 in ED and 1 at home) while the other used a single dose in the ED. The doses and duration of the "control" which was prednisolone were different in each study.

Keller, M. B., et al. (2000). "A comparison of nefazodone, the cognitive behavioral-analysis system of psychotherapy, and their combination for the treatment of chronic depression." *N Engl J Med* 342(20): 1462-1470.

BACKGROUND: Patients with chronic forms of major depression are difficult to treat, and the relative efficacy of medications and psychotherapy is uncertain. **METHODS:** We randomly assigned 681 adults with a chronic nonpsychotic major depressive disorder to 12 weeks of outpatient treatment with nefazodone (maximal dose, 600 mg per day), the cognitive behavioral-analysis system of psychotherapy (16 to 20 sessions), or both. At base line, all patients had scores of at least 20 on the 24-item Hamilton Rating Scale for Depression (indicating clinically significant depression). Remission was defined as a score of 8 or less at weeks 10 and 12. For patients who did not have remission, a satisfactory response was defined as a reduction in the score by at least 50 percent from base line and a score of 15 or less. Raters were unaware of the patients' treatment assignments. **RESULTS:** Of the 681 patients, 662 attended at least one treatment session and were included in the analysis of response. The overall rate of response (both remission and satisfactory response) was 48 percent in both the nefazodone group and in the psychotherapy group, as compared with 73 percent in the combined-treatment group. ($P < 0.001$ for both comparisons). Among the 519 subjects who completed the study, the rates of response were 55 percent in the nefazodone group and 52 percent in the psychotherapy group, as compared with 85 percent in the combined-treatment group ($P < 0.001$ for both comparisons). The rates of withdrawal were similar in the three groups. Adverse events in the nefazodone group were consistent with the known side effects of the drug (e.g., headache, somnolence, dry mouth, nausea, and dizziness). **CONCLUSIONS:** Although about half of patients with chronic forms of major depression have a response to short-term treatment with either nefazodone or a cognitive behavioral-analysis system of psychotherapy, the combination of the two is significantly more efficacious than either treatment alone.

Question: Therapy

Study Type: RCT

Learner Level: Beginner / Intermediate

Notes: Excellent RCT methodology with three randomized arms in this controlled trial. Good discussion points: Good opportunity for discussion of intention to treat. Also clear reporting of results including all reasons for lost to follow up, makes the critical appraisal exercise straightforward. In terms of results, one trickier point is that the primary outcome is a score on the 24-item HRSD scale. Thus, in order to calculate proportion responding, you have to define what change in scale is going to count as a 'satisfactory' therapeutic response. Other good discussion regarding blinding, and attempts to minimize measurement error.

King, T. E., Jr., et al. (2014). "A phase 3 trial of pirfenidone in patients with idiopathic pulmonary fibrosis." *N Engl J Med* 370(22): 2083-2092.

BACKGROUND: In two of three phase 3 trials, pirfenidone, an oral antifibrotic therapy, reduced disease progression, as measured by the decline in forced vital capacity (FVC) or vital capacity, in patients with idiopathic pulmonary fibrosis; in the third trial, this end point was not achieved. We sought to confirm the beneficial effect of pirfenidone on disease progression in such patients. **METHODS:** In this phase 3 study, we randomly assigned 555 patients with idiopathic pulmonary fibrosis to receive either oral pirfenidone (2403 mg per day) or placebo for 52 weeks. The primary end point was the change in FVC or death at week 52. Secondary end points were the 6-minute walk distance, progression-free survival, dyspnea, and death from any cause or from idiopathic pulmonary fibrosis. **RESULTS:** In the pirfenidone group, as compared with the placebo group, there was a relative reduction of 47.9% in the proportion of patients who had an absolute decline of 10 percentage points or more in the percentage of the predicted FVC or who died; there was also a relative increase of 132.5% in the proportion of patients with no decline in FVC ($P < 0.001$). Pirfenidone reduced the decline in the 6-minute walk distance ($P = 0.04$) and improved progression-free survival ($P < 0.001$). There was no significant between-group difference in dyspnea scores ($P = 0.16$) or in rates of death from any cause ($P = 0.10$) or from idiopathic pulmonary fibrosis ($P = 0.23$). However, in a prespecified pooled analysis incorporating results from two previous phase 3 trials, the between-group difference favoring pirfenidone was significant for death from any cause ($P = 0.01$) and from idiopathic pulmonary fibrosis ($P = 0.006$). Gastrointestinal and skin-related adverse events were more common in the pirfenidone group than in the placebo group but rarely led to treatment discontinuation. **CONCLUSIONS:** Pirfenidone, as compared with placebo, reduced disease progression, as reflected by lung function, exercise tolerance, and progression-free survival, in patients with idiopathic pulmonary fibrosis. Treatment was associated with an acceptable side-effect profile and fewer deaths. (Funded by InterMune; ASCEND ClinicalTrials.gov number, NCT01366209.)

Question: Therapy

Study Type: RCT

Learner Level: Beginner

Notes: RCT with good description of methods, good for teaching the difficulties with composite outcomes and role of industry funding.

Konstantinides, S., et al. (2002). "Heparin plus alteplase compared with heparin alone in patients with submassive pulmonary embolism." *N Engl J Med* 347(15): 1143-1150.

BACKGROUND: The use of thrombolytic agents in the treatment of hemodynamically stable patients with acute submassive pulmonary embolism remains controversial. **METHODS:** We conducted a study of patients with acute pulmonary embolism and pulmonary hypertension or right ventricular dysfunction but without arterial hypotension or shock. The patients were randomly assigned in double-blind fashion to receive heparin plus 100 mg of alteplase or heparin plus placebo over a period of two hours. The primary end point was in-hospital death or clinical deterioration requiring an escalation of treatment, which was defined as catecholamine infusion, secondary thrombolysis, endotracheal intubation, cardiopulmonary resuscitation, or emergency surgical embolectomy or thrombus fragmentation by catheter. **RESULTS:** Of 256 patients enrolled, 118 were randomly assigned to receive heparin plus alteplase and 138 to receive heparin plus placebo. The incidence of the primary end point was significantly higher in the heparin-plus-placebo group than in the heparin-plus-alteplase group ($P=0.006$), and the probability of 30-day event-free survival (according to Kaplan-Meier analysis) was higher in the heparin-plus-alteplase group ($P=0.005$). This difference was due to the higher incidence of treatment escalation in the heparin-plus-placebo group (24.6 percent vs. 10.2 percent, $P=0.004$), since mortality was low in both groups (3.4 percent in the heparin-plus-alteplase group and 2.2 percent in the heparin-plus-placebo group, $P=0.71$). Treatment with heparin plus placebo was associated with almost three times the risk of death or treatment escalation that was associated with heparin plus alteplase ($P=0.006$). No fatal bleeding or cerebral bleeding occurred in patients receiving heparin plus alteplase. **CONCLUSIONS:** When given in conjunction with heparin, alteplase can improve the clinical course of stable patients who have acute submassive pulmonary embolism and can prevent clinical deterioration requiring the escalation of treatment during the hospital stay.

Question: Therapy

Study Type: RCT

Learner Level: Intermediate / Advanced

Notes: Prestigious journal and high impact study that has widely been interpreted to justify administration of thrombolytic therapy to relatively stable patients with pulmonary embolism and evidence of right heart strain or failure. It is the only RCT addressing this question. Good discussion points: This is ultimately a negative example. Basic methodology is sound. However there are major problems with the composite outcome, unblinding of the study with respect to one component of the composite and it was also a 'trial stopped early'. A great exercise in unraveling spin in industry sponsored high profile trials.

Kuppermann, N., et al. (2018). "Clinical Trial of Fluid Infusion Rates for Pediatric Diabetic Ketoacidosis." *N Engl J Med* 378(24): 2275-2287.

BACKGROUND: Diabetic ketoacidosis in children may cause brain injuries ranging from mild to severe. Whether intravenous fluids contribute to these injuries has been debated for decades. **METHODS:** We conducted a 13-center, randomized, controlled trial that examined the effects of the rate of administration and the sodium chloride content of intravenous fluids on neurologic outcomes in children with diabetic ketoacidosis. Children were randomly assigned to one of four treatment groups in a 2-by-2 factorial design (0.9% or 0.45% sodium chloride content and rapid or slow rate of administration). The primary outcome was a decline in mental status (two consecutive Glasgow Coma Scale scores of <14 , on a scale ranging from 3 to 15, with lower scores indicating worse mental status) during treatment for diabetic ketoacidosis. Secondary outcomes included clinically apparent brain injury during treatment for diabetic ketoacidosis, short-term memory during treatment for diabetic ketoacidosis, and memory and IQ 2 to 6 months after recovery from diabetic ketoacidosis. **RESULTS:** A total of 1389 episodes of diabetic ketoacidosis were reported in 1255 children. The Glasgow Coma Scale score declined to less than 14 in 48 episodes (3.5%), and clinically apparent brain injury occurred in 12 episodes (0.9%). No significant differences among the treatment groups were observed with respect to the percentage of episodes in which the Glasgow Coma Scale score declined to below 14, the magnitude of decline in the Glasgow Coma Scale score, or the duration of time in which the Glasgow Coma Scale score was less than 14; with respect to the results of the tests of short-term memory; or with respect to the incidence of clinically apparent brain injury during treatment for diabetic ketoacidosis. Memory and IQ scores obtained after the children's recovery from diabetic ketoacidosis also did not differ significantly among the groups. Serious adverse events other than altered mental status were rare and occurred with similar frequency in all treatment groups. **CONCLUSIONS:** Neither the rate of administration nor the sodium chloride content of intravenous fluids significantly influenced neurologic outcomes in children with diabetic ketoacidosis. (Funded by the Eunice Kennedy Shriver National Institute of Child Health and Human Development and the Health Resources and Services Administration; PECARN DKA FLUID ClinicalTrials.gov number, NCT00629707 .).

Question: Therapy

Study Type: RCT

Learner Level: Advanced

Notes: A landmark trial that should change practice. You can discuss study design (2x2 factorial design), patient-oriented outcomes, ITT vs per-protocol analyses, subgroup analyses, type 2 error, should this article change practice or not (and, if not, what more would you need?).

Larson, E. L., et al. (2004). "Effect of antibacterial home cleaning and handwashing products on infectious disease symptoms: a randomized, double-blind trial." *Ann Intern Med* 140(5): 321-329.

BACKGROUND: Despite the widespread household use of cleaning and personal hygiene products containing antibacterial ingredients, their effects on the incidence of infectious disease symptoms have not been studied. **OBJECTIVE:** To evaluate the effect of antibacterial cleaning and handwashing products for consumers on the occurrence of infectious disease symptoms in households.

DESIGN: Randomized, double-blind clinical trial. SETTING: Northern Manhattan inner-city neighborhood, New York. PARTICIPANTS: 238 primarily Hispanic households (1178 persons) that included at least one preschool-age child. Interventions: Households were randomly assigned to use either antibacterial or nonantibacterial products for general cleaning, laundry, and handwashing. All products were commercially available, but the packaging was blinded and the products were provided free to participants. MEASUREMENTS: Hygiene practices and infectious disease symptoms were monitored by weekly telephone calls, monthly home visits, and quarterly interviews for 48 weeks. RESULTS: Symptoms were primarily respiratory: During 26.2% (717 of 2736) of household-months, 23.3% (640 of 2737) of household-months, and 10.2% (278 of 2737) of household-months, one or more members of the household had a runny nose, cough, or sore throat, respectively. Fever was present during 11% (301 of 2737) of household-months, vomiting was present in 2.2% (61 of 2737), diarrhea was present in 2.5% (69 of 2737), and boils or conjunctivitis were present in 0.77% (21 of 2737). Differences between intervention and control groups were not significant for any symptoms (all unadjusted and adjusted relative risks included 1.0) or for numbers of symptoms (overall incidence density ratio, 0.96 [95% CI, 0.82 to 1.12]). CONCLUSIONS: The tested antibacterial products did not reduce the risk for symptoms of viral infectious diseases in households that included essentially healthy persons. This does not preclude the potential contribution of these products to reducing symptoms of bacterial diseases in the home.

Question: Therapy

Study Type: RCT

Learner Level: Beginner / Intermediate

Notes: Fun paper that looked at the effect of antibacterial home cleaning and handwashing products in households with at least one pre-school age child. Fun topic for anyone who has or works with small children (or has to clean their home..) Good discussion points: Nice opportunity to discuss randomization methods looking at the use of antibacterial products as the intervention. Problem with teaching this paper: It is an equivalence trial, therefore you cannot calculate risk reduction or number needed to treat. You can use this as an opportunity to discuss p-values and confidence intervals that are NOT significant.

Le May, S., et al. (2017). "Oral Analgesics Utilization for Children With Musculoskeletal Injury (OUCH Trial): An RCT." *Pediatrics* 140(5).

BACKGROUND: Musculoskeletal injuries (MSK-Is) are a common and painful condition among children that remains poorly treated in the emergency department (ED). We aimed to test the efficacy of a combination of an anti-inflammatory drug with an opioid for pain management of MSK-I in children presenting to the ED. METHODS: In this randomized, double-blinded, placebo-controlled trial, we enrolled children between 6 and 17 years presenting to the ED with an MSK-I and a pain score >29 mm on the visual analog scale (VAS). Participants were randomly assigned to oral morphine (0.2 mg/kg) + ibuprofen (10 mg/kg) (morphine + ibuprofen) or morphine (0.2 mg/kg) + placebo of ibuprofen or ibuprofen (10 mg/kg) + placebo of morphine. Primary outcome was children with VAS pain score <30 mm at 60 minutes postmedication administration. RESULTS: A total of 501 participants were enrolled and 456 were included in primary analyses (morphine + ibuprofen = 177; morphine = 188; ibuprofen = 91). Only 29.9% (morphine + ibuprofen), 29.3% (morphine), and 33.0% (ibuprofen) of participants achieved the primary outcome (P = .81). Mean VAS pain reduction at 60 minutes were -18.7 (95% confidence interval [CI]: -21.9 to -16.6) (morphine + ibuprofen), -17.0 (95% CI: -20.0 to -13.9) (morphine), -18.6 (95% CI: -22.9 to -14.2) (ibuprofen) (P = .69). Children in the morphine + ibuprofen group (P < .001) and in the morphine group (P < .001) experienced more side effects than those in the ibuprofen group. No serious adverse event was reported. CONCLUSIONS: Combination of morphine with ibuprofen did not provide adequate pain relief for children with MSK-I in the ED. None of the study medication provided an optimal pain management because most of children did not reach a mild pain score (NCT02064894).

Question: Therapy

Study Type: RCT

Learner Level: Beginner

Notes: This is an easy article for teaching all of the elements of randomized controlled trials. All elements of validity are touched upon in the study methods. The weaknesses in the study lead to great discussions about intention to treat analysis vs. per protocol and the downstream effects of participants lost to follow up.

Leuppi, J. D., et al. (2013). "Short-term vs conventional glucocorticoid therapy in acute exacerbations of chronic obstructive pulmonary disease: the REDUCE randomized clinical trial." *JAMA* 309(21): 2223-2231.

IMPORTANCE: International guidelines advocate a 7- to 14-day course of systemic glucocorticoid therapy in acute exacerbations of chronic obstructive pulmonary disease (COPD). However, the optimal dose and duration are unknown. OBJECTIVE: To investigate whether a short-term (5 days) systemic glucocorticoid treatment in patients with COPD exacerbation is noninferior to conventional (14 days) treatment in clinical outcome and whether it decreases the exposure to steroids. DESIGN, SETTING, AND PATIENTS REDUCE: (Reduction in the Use of Corticosteroids in Exacerbated COPD), a randomized, noninferiority multicenter trial in 5 Swiss teaching hospitals, enrolling 314 patients presenting to the emergency department with acute COPD exacerbation, past or present smokers (>=20 pack-years) without a history of asthma, from March 2006 through February 2011. INTERVENTIONS: Treatment with 40 mg of prednisone daily for either 5 or 14 days in a placebo-controlled, double-blind fashion. The predefined noninferiority criterion was an absolute increase in exacerbations of at most 15%, translating to a critical hazard ratio of 1.515 for a reference event rate of 50%. MAIN OUTCOME AND MEASURE: Time to next exacerbation within 180 days. RESULTS: Of 314 randomized patients, 289 (92%) of whom were admitted to the hospital, 311 were included in the intention-to-treat analysis and 296 in the per-protocol analysis. Hazard ratios for the short-term vs conventional treatment group were 0.95 (90% CI, 0.70 to 1.29; P = .006 for

noninferiority) in the intention-to-treat analysis and 0.93 (90% CI, 0.68 to 1.26; P = .005 for noninferiority) in the per-protocol analysis, meeting our noninferiority criterion. In the short-term group, 56 patients (35.9%) reached the primary end point; 57 (36.8%) in the conventional group. Estimates of reexacerbation rates within 180 days were 37.2% (95% CI, 29.5% to 44.9%) in the short-term; 38.4% (95% CI, 30.6% to 46.3%) in the conventional, with a difference of -1.2% (95% CI, -12.2% to 9.8%) between the short-term and the conventional. Among patients with a reexacerbation, the median time to event was 43.5 days (interquartile range [IQR], 13 to 118) in the short-term and 29 days (IQR, 16 to 85) in the conventional. There was no difference between groups in time to death, the combined end point of exacerbation, death, or both and recovery of lung function. In the conventional group, mean cumulative prednisone dose was significantly higher (793 mg [95% CI, 710 to 876 mg] vs 379 mg [95% CI, 311 to 446 mg], P < .001), but treatment-associated adverse reactions, including hyperglycemia and hypertension, did not occur more frequently. **CONCLUSIONS AND RELEVANCE:** In patients presenting to the emergency department with acute exacerbations of COPD, 5-day treatment with systemic glucocorticoids was noninferior to 14-day treatment with regard to reexacerbation within 6 months of follow-up but significantly reduced glucocorticoid exposure. These findings support the use of a 5-day glucocorticoid treatment in acute exacerbations of COPD. **TRIAL REGISTRATION:** isrctn.org Identifier: ISRCTN19646069.

Question: Therapy

Study Type: Non-Inferiority RCT

Learner Level: Advanced

Notes: Advanced level, non-inferiority Therapy trial, well reported.

Li, H. K., et al. (2019). "Oral versus Intravenous Antibiotics for Bone and Joint Infection." *N Engl J Med* 380(5): 425-436.

BACKGROUND: The management of complex orthopedic infections usually includes a prolonged course of intravenous antibiotic agents. We investigated whether oral antibiotic therapy is noninferior to intravenous antibiotic therapy for this indication. **METHODS:** We enrolled adults who were being treated for bone or joint infection at 26 U.K. centers. Within 7 days after surgery (or, if the infection was being managed without surgery, within 7 days after the start of antibiotic treatment), participants were randomly assigned to receive either intravenous or oral antibiotics to complete the first 6 weeks of therapy. Follow-on oral antibiotics were permitted in both groups. The primary end point was definitive treatment failure within 1 year after randomization. In the analysis of the risk of the primary end point, the noninferiority margin was 7.5 percentage points. **RESULTS:** Among the 1054 participants (527 in each group), end-point data were available for 1015 (96.3%). Treatment failure occurred in 74 of 506 participants (14.6%) in the intravenous group and 67 of 509 participants (13.2%) in the oral group. Missing end-point data (39 participants, 3.7%) were imputed. The intention-to-treat analysis showed a difference in the risk of definitive treatment failure (oral group vs. intravenous group) of -1.4 percentage points (90% confidence interval [CI], -4.9 to 2.2; 95% CI, -5.6 to 2.9), indicating noninferiority. Complete-case, per-protocol, and sensitivity analyses supported this result. The between-group difference in the incidence of serious adverse events was not significant (146 of 527 participants [27.7%] in the intravenous group and 138 of 527 [26.2%] in the oral group; P=0.58). Catheter complications, analyzed as a secondary end point, were more common in the intravenous group (9.4% vs. 1.0%). **CONCLUSIONS:** Oral antibiotic therapy was noninferior to intravenous antibiotic therapy when used during the first 6 weeks for complex orthopedic infection, as assessed by treatment failure at 1 year. (Funded by the National Institute for Health Research; OVIVA Current Controlled Trials number, ISRCTN91566927 .).

Question: Therapy

Study Type: RCT

Learner Level: Intermediate

Notes: Good article to teach non-inferiority.

Liu, Z., et al. (2009). "Carotid artery stenting versus carotid endarterectomy: systematic review and meta-analysis." *World J Surg* 33(3): 586-596.

BACKGROUND: The aim of this study was to evaluate and investigate the complications of carotid endarterectomy (CEA) and carotid artery stenting (CAS) by performing a meta-analysis based on prospective randomized controlled trials (RCTs). **METHODS:** We performed a search of multiple electronic databases for RCTs containing patients with carotid stenosis who underwent CAS or CEA, focusing on studies published during 1995-2008. **RESULTS:** Eight trials with 2942 patients (1462 with CEA, 1480 with CAS) were analyzed. The pooled relative risk (RR) after CEA for stroke/death 30 days or 1 year was similar to that for CAS. Thirty-day RR = 0.69, 95% confidence interval (CI) = 0.45-1.07, p = 0.10. One-year RR = 0.88, 95% CI = 0.43-1.79, p = 0.72. The rates of death, disabling stroke, and nondisabling stroke at 30 days did not differ significantly between CEA and CAS in the subgroup analysis. Compared with CEA, the relative risk of disabling stroke/death within 30 days was not significantly less for CAS with embolic protection devices (EPDs). The relative risk of myocardial infarction within 30 days, myocardial infarction within 1 year, and cervical/peripheral nerve injury within 30 days were significantly higher after CEA; the relative risk of bradycardia/hypotension within 30 days and the 1-year restenosis rate were significantly higher after CAS. **CONCLUSIONS:** CAS is equal to CEA with regard to the incidence of stroke/death. These procedures may be considered complementary rather than competing modes of therapy, each of which can be optimized with careful patient selection. CAS with an EPD may be appropriate in certain patients, and in general CAS should be considered cautiously in symptomatic patients.

Question: Therapy

Study Type: Meta-analysis

Learner Level: Intermediate

Notes: Good to teach heterogeneity

Maitland, K., et al. (2011). "Mortality after fluid bolus in African children with severe infection." *N Engl J Med* 364(26): 2483-2495.

BACKGROUND: The role of fluid resuscitation in the treatment of children with shock and life-threatening infections who live in resource-limited settings is not established. **METHODS:** We randomly assigned children with severe febrile illness and impaired perfusion to receive boluses of 20 to 40 ml of 5% albumin solution (albumin-bolus group) or 0.9% saline solution (saline-bolus group) per kilogram of body weight or no bolus (control group) at the time of admission to a hospital in Uganda, Kenya, or Tanzania (stratum A); children with severe hypotension were randomly assigned to one of the bolus groups only (stratum B). All children received appropriate antimicrobial treatment, intravenous maintenance fluids, and supportive care, according to guidelines. Children with malnutrition or gastroenteritis were excluded. The primary end point was 48-hour mortality; secondary end points included pulmonary edema, increased intracranial pressure, and mortality or neurologic sequelae at 4 weeks. **RESULTS:** The data and safety monitoring committee recommended halting recruitment after 3141 of the projected 3600 children in stratum A were enrolled. Malaria status (57% overall) and clinical severity were similar across groups. The 48-hour mortality was 10.6% (111 of 1050 children), 10.5% (110 of 1047 children), and 7.3% (76 of 1044 children) in the albumin-bolus, saline-bolus, and control groups, respectively (relative risk for saline bolus vs. control, 1.44; 95% confidence interval [CI], 1.09 to 1.90; $P=0.01$; relative risk for albumin bolus vs. saline bolus, 1.01; 95% CI, 0.78 to 1.29; $P=0.96$; and relative risk for any bolus vs. control, 1.45; 95% CI, 1.13 to 1.86; $P=0.003$). The 4-week mortality was 12.2%, 12.0%, and 8.7% in the three groups, respectively ($P=0.004$ for the comparison of bolus with control). Neurologic sequelae occurred in 2.2%, 1.9%, and 2.0% of the children in the respective groups ($P=0.92$), and pulmonary edema or increased intracranial pressure occurred in 2.6%, 2.2%, and 1.7% ($P=0.17$), respectively. In stratum B, 69% of the children (9 of 13) in the albumin-bolus group and 56% (9 of 16) in the saline-bolus group died ($P=0.45$). The results were consistent across centers and across subgroups according to the severity of shock and status with respect to malaria, coma, sepsis, acidosis, and severe anemia. **CONCLUSIONS:** Fluid boluses significantly increased 48-hour mortality in critically ill children with impaired perfusion in these resource-limited settings in Africa. (Funded by the Medical Research Council, United Kingdom; FEAST Current Controlled Trials number, ISRCTN69856593.).

Question: Therapy

Study Type: RCT

Learner Level: Intermediate/Advanced

Notes: I just used this paper for the Capstone EBM course, and it seemed to work well as a teaching paper. Good methods. Doesn't have blinding, but that leads to good discussion about situations when you can't blind. Also, there were worse outcomes in the intervention group, so you end up calculating a RRI and NNH. The results of this paper go against standard practice, so it generates good discussion of why that might have occurred.

March, J., et al. (2004). "Fluoxetine, cognitive-behavioral therapy, and their combination for adolescents with depression: Treatment for Adolescents With Depression Study (TADS) randomized controlled trial." *JAMA* 292(7): 807-820.

CONTEXT: Initial treatment of major depressive disorder in adolescents may include cognitive-behavioral therapy (CBT) or a selective serotonin reuptake inhibitor (SSRI). However, little is known about their relative or combined effectiveness. **OBJECTIVE:** To evaluate the effectiveness of 4 treatments among adolescents with major depressive disorder. **DESIGN, SETTING, AND PARTICIPANTS:** Randomized controlled trial of a volunteer sample of 439 patients between the ages of 12 to 17 years with a primary Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition, diagnosis of major depressive disorder. The trial was conducted at 13 US academic and community clinics between spring 2000 and summer 2003. **INTERVENTIONS:** Twelve weeks of (1) fluoxetine alone (10 to 40 mg/d), (2) CBT alone, (3) CBT with fluoxetine (10 to 40 mg/d), or (4) placebo (equivalent to 10 to 40 mg/d). Placebo and fluoxetine alone were administered double-blind; CBT alone and CBT with fluoxetine were administered unblinded. **MAIN OUTCOME MEASURES:** Children's Depression Rating Scale-Revised total score and, for responder analysis, a (dichotomized) Clinical Global Impressions improvement score. **RESULTS:** Compared with placebo, the combination of fluoxetine with CBT was statistically significant ($P=.001$) on the Children's Depression Rating Scale-Revised. Compared with fluoxetine alone ($P=.02$) and CBT alone ($P=.01$), treatment of fluoxetine with CBT was superior. Fluoxetine alone is a superior treatment to CBT alone ($P=.01$). Rates of response for fluoxetine with CBT were 71.0% (95% confidence interval [CI], 62%-80%); fluoxetine alone, 60.6% (95% CI, 51%-70%); CBT alone, 43.2% (95% CI, 34%-52%); and placebo, 34.8% (95% CI, 26%-44%). On the Clinical Global Impressions improvement responder analysis, the 2 fluoxetine-containing conditions were statistically superior to CBT and to placebo. Clinically significant suicidal thinking, which was present in 29% of the sample at baseline, improved significantly in all 4 treatment groups. Fluoxetine with CBT showed the greatest reduction ($P=.02$). Seven (1.6%) of 439 patients attempted suicide; there were no completed suicides. **CONCLUSION:** The combination of fluoxetine with CBT offered the most favorable tradeoff between benefit and risk for adolescents with major depressive disorder.

Question: Therapy

Study Type: RCT

Learner Level: Beginner / Intermediate

Notes: This Multicenter RCT is a landmark trial in adolescents as the first large trial to compare placebo to cognitive behavior +/- antidepressant medication (Fluoxetine) in this age group. Strong, clearly reported methods; Evidence-based Mental Health Summary 2005; 8: 10; Possible Discussion Points: The trial is also notable for a timely publication given that in February 2004 the FDA convened an advisory panel to review possible increased risk of suicidality in patients on SSRI medications. There was intense media spotlight associated with the FDA black box warning describing potential increase risk of suicidal thoughts and behaviors in children and adolescents. (See related article in table: Antidepressant treatment and the risk of fatal and non-fatal self harm in first episode depression; nested case-control study, Harm)

Martinez, F. J., et al. (2005). "The clinical course of patients with idiopathic pulmonary fibrosis." *Ann Intern Med* 142(12 Pt 1): 963-967.

BACKGROUND: Prospective data defining the clinical course in idiopathic pulmonary fibrosis (IPF) are sparse. **OBJECTIVE:** To analyze the clinical course of patients with mild to moderate IPF. **DESIGN:** Analysis of data from the placebo group of a randomized, controlled trial evaluating interferon-gamma1b. **SETTING:** Academic and community medical centers. **PATIENTS:** 168 patients in the placebo group of a trial evaluating interferon-gamma1b. **MEASUREMENTS:** Measures of physiology and dyspnea assessed at 12-week intervals; hospitalizations; and the pace of deterioration and cause of death over a median period of 76 weeks. **RESULTS:** Physiologic variables changed minimally during the study. However, 23% of patients required hospitalization for a respiratory disorder and 21% died. Idiopathic pulmonary fibrosis was the primary cause of death in 89% of patients who died, and an apparent acute clinical deterioration preceded death in 47% of these patients. **LIMITATIONS:** The instrument used to define the pace of deterioration and cause of death was applied retrospectively. **CONCLUSIONS:** Recognition of the common occurrence of acute fatal deterioration in patients with mild to moderate IPF has important implications for monitoring patients and supports early referral for lung transplantation.

Question: Therapy

Study Type: RCT

Learner Level: Beginner/Intermediate

Notes: Solid Methodology; Landmark paper in internal medicine; Important and controversial clinical question (Dig: to treat or not to treat?); Good discussion points: Can discuss both equivalency (outcomes that did not show difference) and also positive findings (can calculate an NNT- number needed to treat); Can discuss the balance of benefits of therapy vs. potential harms; Good clinical applicability discussion can follow; Does a 'good' paper ever grow old? (i.e. does publication date matter?).. if so, when?

Marx, C. E., et al. (2014). "Proof-of-concept randomized controlled trial of pregnenolone in schizophrenia." *Psychopharmacology* 231(17): 3647-3662.

RATIONALE: Preclinical and clinical data suggest that pregnenolone may be a promising therapeutic in schizophrenia. Pregnenolone is neuroprotective and enhances learning and memory, myelination, and microtubule polymerization. Treatment with pregnenolone elevates allopregnanolone (a neurosteroid that enhances GABAA receptor responses) and pregnenolone sulfate (a positive NMDA receptor modulator). Pregnenolone could thus potentially mitigate GABA dysregulation and/or NMDA receptor hypofunction in schizophrenia via metabolism to other neurosteroids. **OBJECTIVE:** The objective of this study is to conduct a randomized controlled trial of adjunctive pregnenolone in schizophrenia. **METHODS:** Following a placebo lead-in, 120 participants were randomized to pregnenolone or placebo for 8 weeks (Institute for Mental Health, Singapore). Primary endpoints were changes in MATRICS Consensus Cognitive Battery (MCCB) composite scores (cognitive symptoms), UCSD Performance-based Skills Assessment-Brief (UPSA-B) composite scores (functional capacity), and Scale for Assessment of Negative Symptoms (SANS) total scores (negative symptoms). A modified intent-to-treat analysis approach was utilized. **RESULTS:** No significant changes compared to placebo were demonstrated in composite MCCB scores. In contrast, participants randomized to pregnenolone (n = 56) demonstrated greater improvements in functional capacity (UPSA-B composite changes) compared to placebo (n = 55), p = 0.03. Pregnenolone was also superior to placebo in the communication subscale of the UPSA-B (p < 0.001). Serum pregnenolone changes post-treatment were correlated with UPSA-B composite score changes in females (r s = 0.497, p < 0.042, n = 17) but not in males. Mean total SANS scores were very low at baseline and did not improve further post-treatment. Pregnenolone was well-tolerated. **CONCLUSIONS:** Pregnenolone improved functional capacity in participants with schizophrenia, but did not improve cognitive symptoms over an 8-week treatment period. Neurosteroid changes correlated with functional improvements in female participants. Neurosteroid interventions may exhibit promise as new therapeutic leads for schizophrenia.

Question: Therapy

Study Type: RCT

Learner Level: Intermediate

Notes: Neurosteroids are hot topics in mental health as is the search for drugs that may improve cognition in schizophrenia. This is a great teaching paper because it's timely and clinically relevant and written by a Duke author! It challenges learners to evaluate a potential new therapy and decide if the evidence is sufficiently strong to change their clinical practice. The paper itself illustrates a number of key EBM concepts: the limitations of small sample sizes, identifying unreported prognostic factors that may have affected outcome, a priori vs. post-hoc analyses, corrections for multiple comparisons, generalizability to our patients, calculating an effect size from the reported results measured in raw scores, etc...

McAlindon, T. E., et al. (2017). "Effect of Intra-articular Triamcinolone vs Saline on Knee Cartilage Volume and Pain in Patients With Knee Osteoarthritis: A Randomized Clinical Trial." *JAMA* 317(19): 1967-1975.

Importance: Synovitis is common and is associated with progression of structural characteristics of knee osteoarthritis. Intra-articular corticosteroids could reduce cartilage damage associated with synovitis but might have adverse effects on cartilage and periarticular bone. Objective: To determine the effects of intra-articular injection of 40 mg of triamcinolone acetonide every 3 months on progression of cartilage loss and knee pain. Design, Setting, and Participants: Two-year, randomized, placebo-controlled, double-blind trial of intra-articular triamcinolone vs saline for symptomatic knee osteoarthritis with ultrasonic features of synovitis in 140 patients. Mixed-effects regression models with a random intercept were used to analyze the longitudinal repeated outcome measures. Patients fulfilling the American College of Rheumatology criteria for symptomatic knee osteoarthritis, Kellgren-Lawrence grades 2 or 3, were enrolled at Tufts Medical Center beginning February 11, 2013; all patients completed the study by January 1, 2015. Interventions: Intra-articular triamcinolone (n = 70) or saline (n = 70) every 12 weeks for 2 years. Main Outcomes and Measures: Annual knee magnetic resonance imaging for quantitative evaluation of cartilage volume (minimal clinically important difference not yet defined), and Western Ontario and McMaster Universities Osteoarthritis index collected every 3 months (Likert pain subscale range, 0 [no pain] to 20 [extreme pain]; minimal clinically important improvement, 3.94). Results: Among 140 randomized patients (mean age, 58 [SD, 8] years, 75 women [54%]), 119 (85%) completed the study. Intra-articular triamcinolone resulted in significantly greater cartilage volume loss than did saline for a mean change in index compartment cartilage thickness of -0.21 mm vs -0.10 mm (between-group difference, -0.11 mm; 95% CI, -0.20 to -0.03 mm); and no significant difference in pain (-1.2 vs -1.9; between-group difference, -0.6; 95% CI, -1.6 to 0.3). The saline group had 3 treatment-related adverse events compared with 5 in the triamcinolone group and had a small increase in hemoglobin A1c levels (between-group difference, -0.2%; 95% CI, -0.5% to -0.007%). Conclusions and Relevance: Among patients with symptomatic knee osteoarthritis, 2 years of intra-articular triamcinolone, compared with intra-articular saline, resulted in significantly greater cartilage volume loss and no significant difference in knee pain. These findings do not support this treatment for patients with symptomatic knee osteoarthritis. Trial Registration: ClinicalTrials.gov Identifier: NCT01230424.

Question: Therapy

Study Type: RCT

Learner Level: Beginner

Notes: A good article for discussing the application of results to specific patients' questions (i.e., the steroids may not do much long term, but are we sure they won't help the patient get to rehab?).

McFalls, E. O., et al. (2004). "Coronary-artery revascularization before elective major vascular surgery." *N Engl J Med* 351(27): 2795-2804.

BACKGROUND: The benefit of coronary-artery revascularization before elective major vascular surgery is unclear. **METHODS:** We randomly assigned patients at increased risk for perioperative cardiac complications and clinically significant coronary artery disease to undergo either revascularization or no revascularization before elective major vascular surgery. The primary end point was long-term mortality. **RESULTS:** Of 5859 patients scheduled for vascular operations at 18 Veterans Affairs medical centers, 510 (9 percent) were eligible for the study and were randomly assigned to either coronary-artery revascularization before surgery or no revascularization before surgery. The indications for a vascular operation were an expanding abdominal aortic aneurysm (33 percent) or arterial occlusive disease of the legs (67 percent). Among the patients assigned to preoperative coronary-artery revascularization, percutaneous coronary intervention was performed in 59 percent, and bypass surgery was performed in 41 percent. The median time from randomization to vascular surgery was 54 days in the revascularization group and 18 days in the group not undergoing revascularization ($P < 0.001$). At 2.7 years after randomization, mortality in the revascularization group was 22 percent and in the no-revascularization group 23 percent (relative risk, 0.98; 95 percent confidence interval, 0.70 to 1.37; $P = 0.92$). Within 30 days after the vascular operation, a postoperative myocardial infarction, defined by elevated troponin levels, occurred in 12 percent of the revascularization group and 14 percent of the no-revascularization group ($P = 0.37$). **CONCLUSIONS:** Coronary-artery revascularization before elective vascular surgery does not significantly alter the long-term outcome. On the basis of these data, a strategy of coronary-artery revascularization before elective vascular surgery among patients with stable cardiac symptoms cannot be recommended.

Question: Therapy

Study Type: RCT

Learner Level: Intermediate

Notes: RCT of patients at risk for perioperative cardiac complications assigned to revascularization or no revascularization prior to elective major vascular surgery. landmark trial that changed thinking about perioperative cardiac risk assessment and intervention. Good paper for discussion of methods and incorporation of evidence into practice. Because the results are 'no difference' between groups, you can not calculate an NNT.

McMahon, S. R., et al. (1997). "Parents can dose liquid medication accurately." *Pediatrics* 100(3 Pt 1): 330-333.

OBJECTIVE: To determine whether parental errors in dosing liquid medication can be decreased through education. **DESIGN:** Randomized convenience sample stratified to three study groups. **SETTING:** General pediatric clinic, largely indigent and Latino. **PATIENTS:** A total of 45 English-speaking and 45 Spanish-speaking children diagnosed with otitis media and treated with an antibiotic suspension. **INTERVENTION:** Group 1 patients received the prescription and verbal instructions. Group 2 patients received the prescription and a syringe, then the correct dose was demonstrated. Group 3 patients received the prescription, a syringe with a line marked at the correct dose, and a demonstration. After returning from the pharmacy, parents administered the medication under

observation. Parents in group 1 used a dispensing device similar to that planned for home use. The other groups used the syringe. After observation but before discharge, everyone received a syringe with a line marked at the correct dose. Patients were seen again at approximately 1 month, and parents demonstrated how much medication they had administered. **MAIN OUTCOME MEASURE:** Percent of parents who administered the correct dose. **RESULTS:** Patients in group 1 received between 32% and 147% of the correct dose, with only 11 of 30 (37%) receiving the correct dose (+/-0.2 mL). In group 2, 25 of 30 (83%) parents administered the correct dose, and in group 3, 30 of 30 (100%) gave the correct dose. Simultaneous logistic regression indicated that accuracy of dosage differed across instructional groups and language. At follow-up, 23 of 26 parents demonstrated the correct dose. **CONCLUSION:** Education can decrease medication dosing errors made by both Spanish-speaking and English-speaking parents. Effectiveness was also shown at follow-up.

Question: Therapy

Study Type: RCT

Learner Level: Beginner / Intermediate

Notes: Good RCT about an intervention other than a drug. Reasonably good methodology. Largely ignored paper that could have important implications. Discussion points: Role of complete follow-up. Role of concealment of randomization and allocation can be discussed. Easy calculations of NNT. Can change management strategy for pediatricians!

Miller, E. R., 3rd, et al. (2005). "Meta-analysis: high-dosage vitamin E supplementation may increase all-cause mortality." *Ann Intern Med* 142(1): 37-46.

BACKGROUND: Experimental models and observational studies suggest that vitamin E supplementation may prevent cardiovascular disease and cancer. However, several trials of high-dosage vitamin E supplementation showed non-statistically significant increases in total mortality. **PURPOSE:** To perform a meta-analysis of the dose-response relationship between vitamin E supplementation and total mortality by using data from randomized, controlled trials. **PATIENTS:** 135,967 participants in 19 clinical trials. Of these trials, 9 tested vitamin E alone and 10 tested vitamin E combined with other vitamins or minerals. The dosages of vitamin E ranged from 16.5 to 2000 IU/d (median, 400 IU/d). **DATA SOURCES:** PubMed search from 1966 through August 2004, complemented by a search of the Cochrane Clinical Trials Database and review of citations of published reviews and meta-analyses. No language restrictions were applied. **DATA EXTRACTION:** 3 investigators independently abstracted study reports. The investigators of the original publications were contacted if required information was not available. **DATA SYNTHESIS:** 9 of 11 trials testing high-dosage vitamin E (> or =400 IU/d) showed increased risk (risk difference > 0) for all-cause mortality in comparisons of vitamin E versus control. The pooled all-cause mortality risk difference in high-dosage vitamin E trials was 39 per 10,000 persons (95% CI, 3 to 74 per 10,000 persons; P = 0.035). For low-dosage vitamin E trials, the risk difference was -16 per 10,000 persons (CI, -41 to 10 per 10,000 persons; P > 0.2). A dose-response analysis showed a statistically significant relationship between vitamin E dosage and all-cause mortality, with increased risk of dosages greater than 150 IU/d. **LIMITATIONS:** High-dosage (> or =400 IU/d) trials were often small and were performed in patients with chronic diseases. The generalizability of the findings to healthy adults is uncertain. Precise estimation of the threshold at which risk increases is difficult. **CONCLUSION:** High-dosage (> or =400 IU/d) vitamin E supplements may increase all-cause mortality and should be avoided.

Question: Therapy

Study Type: Meta-analysis

Learner Level: Intermediate

Notes: Solid methodology; Recent, pertinent to many disciplines; Good Discussion Points: Nice illustration of how meta-analysis is used to detect important, but rare adverse events occurring in clinical trials. Results are clearly laid out including data table, forest plots, dose-effect curve; Good paper to discuss confidence intervals (Figures 2 and 4 on pages 42 and 44). The length of the paper may intimidate some earlier learners

Moher, D., et al. (1999). "Improving the quality of reports of meta-analyses of randomised controlled trials: the QUOROM statement. Quality of Reporting of Meta-analyses." *Lancet* 354(9193): 1896-1900.

BACKGROUND: The Quality of Reporting of Meta-analyses (QUOROM) conference was convened to address standards for improving the quality of reporting of meta-analyses of clinical randomised controlled trials (RCTs). **METHODS:** The QUOROM group consisted of 30 clinical epidemiologists, clinicians, statisticians, editors, and researchers. In conference, the group was asked to identify items they thought should be included in a checklist of standards. Whenever possible, checklist items were guided by research evidence suggesting that failure to adhere to the item proposed could lead to biased results. A modified Delphi technique was used in assessing candidate items. **FINDINGS:** The conference resulted in the QUOROM statement, a checklist, and a flow diagram. The checklist describes our preferred way to present the abstract, introduction, methods, results, and discussion sections of a report of a meta-analysis. It is organised into 21 headings and subheadings regarding searches, selection, validity assessment, data abstraction, study characteristics, and quantitative data synthesis, and in the results with "trial flow", study characteristics, and quantitative data synthesis; research documentation was identified for eight of the 18 items. The flow diagram provides information about both the numbers of RCTs identified, included, and excluded and the reasons for exclusion of trials. **INTERPRETATION:** We hope this report will generate further thought about ways to improve the quality of reports of meta-analyses of RCTs and that interested readers, reviewers, researchers, and editors will use the QUOROM statement and generate ideas for its improvement.

Question: Therapy

Study Type: Meta-analysis

Learner Level: Intermediate

Notes: Important question which prompts good discussion of risks v. benefits v. cost. Good Discussion Points: Strengths: Well designed with clear description of methods. Can use methods to prompt vocabulary discussion for concepts pertaining to Meta-analysis (e.g. methods for validity assessment, summary treatment effects, sensitivity analysis, heterogeneity); Weaknesses: Variety of interventions can become very confusing to learners. More Advanced Discussion: might consider getting the QUOROM statement that is referenced in the methods of the paper (p. 859) for more in depth discussion of Meta-analysis methodology. Check list on p 1897 can be used to assess this (or any other systematic review); See also Hull 2001.

Mohiuddin, S. M., et al. (2007). "Intensive smoking cessation intervention reduces mortality in high-risk smokers with cardiovascular disease." *Chest* 131(2): 446-452.

PURPOSES: To compare an intensive smoking cessation intervention against usual care in hospitalized high-risk smokers with acute cardiovascular disease. **METHODS:** A total of 209 hospitalized smokers were randomized to the intensive intervention (n = 109) or to usual care (n = 100). Usual care consisted only of counseling and printed educational material provided prior to hospital discharge. Intensive treatment consisted of a minimum of 12 weeks of behavior modification counseling and individualized pharmacotherapy provided at no cost to the participant. Smoking status in all subjects was confirmed biochemically (ie, by measuring expired carbon monoxide) at 3, 6, 12, and 24 months after randomization. Outcomes included point prevalence and continuous abstinence smoking cessation rates, hospitalizations, and all-cause mortality. **RESULTS:** At each follow-up interval, point prevalence and continuous abstinence smoking cessation rates were significantly greater in the intensive-treatment group compared to the usual-care group. At 24 months, continuous abstinence smoking cessation rates were 33% in the intensive-treatment group and 9% in the usual-care group (p < 0.0001). Over the 2-year follow-up period, 41 patients in the usual-care group were hospitalized compared to 25 patients in the intensive-treatment group (relative risk reduction [RRR], 44%; 95% confidence interval [CI], 16 to 63%; p = 0.007). The all-cause mortality rate was 2.8% in the intensive-treatment group and 12.0% in the usual-care group (RRR, 77%; 95% CI, 27 to 93%; p = 0.014). The absolute risk reduction in mortality was 9.2% with a number needed to treat of 11. **CONCLUSION:** Hospitalized smokers, especially those with cardiovascular disease, should undergo treatment with a structured intensive cessation intervention. The duration of the initial treatment should be 3 months.

Question: Therapy

Study Type: RCT

Learner Level: .

Notes: Therapy

Molyneux, A., et al. (2003). "Clinical trial comparing nicotine replacement therapy (NRT) plus brief counselling, brief counselling alone, and minimal intervention on smoking cessation in hospital inpatients." *Thorax* 58(6): 484-488.

BACKGROUND: Guidelines recommend that smoking cessation interventions are offered in all clinical settings to all smokers willing to make a quit attempt. Since the effectiveness of routine provision of behavioural counselling and nicotine replacement therapy (NRT) to smokers admitted to hospital has not been established, a randomised controlled trial of these interventions given together compared with counselling alone or minimal intervention was performed in hospital inpatients. **METHODS:** Medical and surgical inpatients who were current smokers at the time of admission were randomised to receive either usual care (no additional advice at admission), counselling alone (20 minute intervention with written materials), or NRT plus counselling (counselling intervention with a 6 week course of NRT). Continuous and point prevalence abstinence from smoking (validated by exhaled carbon monoxide <10 ppm) was measured at discharge from hospital and at 3 and 12 months, and self-reported reduction in cigarette consumption in smokers was assessed at 3 and 12 months. **RESULTS:** 274 inpatient smokers were enrolled. Abstinence was higher in the NRT plus counselling group (n=91) than in the counselling alone (n=91) or usual care (n=92) groups. The difference between the groups was significant for validated point prevalence abstinence at discharge (55%, 43%, 37% respectively, p=0.045) and at 12 months (17%, 6%, 8%, p=0.03). The respective differences in continuous validated abstinence at 12 months were 11%, 4%, 8% (p=0.25). There was no significant difference between counselling alone and usual care, or in reduction in cigarette consumption between the treatment groups. **CONCLUSIONS:** NRT given with brief counselling to hospital inpatients is an effective routine smoking cessation intervention.

Question: Therapy

Study Type: RCT

Learner Level: .

Notes: Therapy

Nichol, G., et al. (2015). "Trial of Continuous or Interrupted Chest Compressions during CPR." *N Engl J Med* 373(23): 2203-2214.

BACKGROUND: During cardiopulmonary resuscitation (CPR) in patients with out-of-hospital cardiac arrest, the interruption of manual chest compressions for rescue breathing reduces blood flow and possibly survival. We assessed whether outcomes after continuous compressions with positive-pressure ventilation differed from those after compressions that were interrupted for

ventilations at a ratio of 30 compressions to two ventilations. **METHODS:** This cluster-randomized trial with crossover included 114 emergency medical service (EMS) agencies. Adults with non-trauma-related cardiac arrest who were treated by EMS providers received continuous chest compressions (intervention group) or interrupted chest compressions (control group). The primary outcome was the rate of survival to hospital discharge. Secondary outcomes included the modified Rankin scale score (on a scale from 0 to 6, with a score of ≤ 3 indicating favorable neurologic function). CPR process was measured to assess compliance. **RESULTS:** Of 23,711 patients included in the primary analysis, 12,653 were assigned to the intervention group and 11,058 to the control group. A total of 1129 of 12,613 patients with available data (9.0%) in the intervention group and 1072 of 11,035 with available data (9.7%) in the control group survived until discharge (difference, -0.7 percentage points; 95% confidence interval [CI], -1.5 to 0.1; $P=0.07$); 7.0% of the patients in the intervention group and 7.7% of those in the control group survived with favorable neurologic function at discharge (difference, -0.6 percentage points; 95% CI, -1.4 to 0.1, $P=0.09$). Hospital-free survival was significantly shorter in the intervention group than in the control group (mean difference, -0.2 days; 95% CI, -0.3 to -0.1; $P=0.004$). **CONCLUSIONS:** In patients with out-of-hospital cardiac arrest, continuous chest compressions during CPR performed by EMS providers did not result in significantly higher rates of survival or favorable neurologic function than did interrupted chest compressions. (Funded by the National Heart, Lung, and Blood Institute and others; ROC CCC ClinicalTrials.gov number, NCT01372748.)

Question: Therapy

Study Type: RCT

Learner Level: Intermediate/Advanced

Notes: Intermediate to advanced – good for discussion on per protocol vs. intention-to-treat analysis.

Olasveengen, T. M., et al. (2009). "Intravenous drug administration during out-of-hospital cardiac arrest: a randomized trial." *JAMA* 302(20): 2222-2229.

CONTEXT: Intravenous access and drug administration are included in advanced cardiac life support (ACLS) guidelines despite a lack of evidence for improved outcomes. Epinephrine was an independent predictor of poor outcome in a large epidemiological study, possibly due to toxicity of the drug or cardiopulmonary resuscitation (CPR) interruptions secondary to establishing an intravenous line and drug administration. **OBJECTIVE:** To determine whether removing intravenous drug administration from an ACLS protocol would improve survival to hospital discharge after out-of-hospital cardiac arrest. **DESIGN, SETTING, AND PATIENTS:** Prospective, randomized controlled trial of consecutive adult patients with out-of-hospital nontraumatic cardiac arrest treated within the emergency medical service system in Oslo, Norway, between May 1, 2003, and April 28, 2008. **INTERVENTIONS:** Advanced cardiac life support with intravenous drug administration or ACLS without access to intravenous drug administration. **MAIN OUTCOME MEASURES:** The primary outcome was survival to hospital discharge. The secondary outcomes were 1-year survival, survival with favorable neurological outcome, hospital admission with return of spontaneous circulation, and quality of CPR (chest compression rate, pauses, and ventilation rate). **RESULTS:** Of 1183 patients for whom resuscitation was attempted, 851 were included; 418 patients were in the ACLS with intravenous drug administration group and 433 were in the ACLS with no access to intravenous drug administration group. The rate of survival to hospital discharge was 10.5% for the intravenous drug administration group and 9.2% for the no intravenous drug administration group ($P = .61$), 32% vs 21%, respectively, ($P < .001$) for hospital admission with return of spontaneous circulation, 9.8% vs 8.1% ($P = .45$) for survival with favorable neurological outcome, and 10% vs 8% ($P = .53$) for survival at 1 year. The quality of CPR was comparable and within guideline recommendations for both groups. After adjustment for ventricular fibrillation, response interval, witnessed arrest, or arrest in a public location, there was no significant difference in survival to hospital discharge for the intravenous group vs the no intravenous group (adjusted odds ratio, 1.15; 95% confidence interval, 0.69-1.91). **CONCLUSION:** Compared with patients who received ACLS without intravenous drug administration following out-of-hospital cardiac arrest, patients with intravenous access and drug administration had higher rates of short-term survival with no statistically significant improvement in survival to hospital discharge, quality of CPR, or long-term survival. **TRIAL REGISTRATION:** clinicaltrials.gov Identifier: NCT00121524.

Question: Therapy

Study Type: RCT

Learner Level: Beginner - Intermediate

Notes: Controversial topic- good ethics discussion about ethics of randomization; Very good for discussion of allocation concealment and intention to treat analysis. Not good for teaching therapy math because there are no numbers amenable for NNT or NNH calculation

Oler, A., et al. (1996). "Adding heparin to aspirin reduces the incidence of myocardial infarction and death in patients with unstable angina. A meta-analysis." *JAMA* 276(10): 811-815.

OBJECTIVE: To estimate the risk of myocardial infarction (MI) and death in patients with unstable angina who are treated with aspirin plus heparin compared with patients treated with aspirin alone. **DATA SOURCES:** Studies were retrieved using MEDLINE, bibliographies, and consultation with experts. **STUDY SELECTION:** Only published trials that enrolled patients with unstable angina, randomized participants to aspirin plus heparin vs aspirin alone, and reported incidence of myocardial infarction or death were included in the meta-analysis. **DATA EXTRACTION:** Patient outcomes including MI or death, recurrent ischemic pain, and major bleeding during randomized treatment; revascularization procedures after randomization; and MI or death during the 2 to 12 weeks following randomization were extracted by 2 authors, 1 of whom was blinded to the journal, institution, and author of each study.

DATA SYNTHESIS: Six randomized trials were included. The overall summary relative risk (RR) of MI or death during randomized treatment was 0.67 (95% confidence interval [CI], 0.44-1.02) in patients with unstable angina treated with aspirin plus heparin compared with those treated with aspirin alone. The summary RRs for secondary endpoints in patients treated with aspirin plus heparin compared with those treated with aspirin alone were 0.68 (95% CI, 0.40-1.17) for recurrent ischemic pain; 0.82 (95% CI, 0.56-1.20) for MI or death 2 to 12 weeks following randomization; 1.03 (95% CI, 0.74-1.43) for revascularization; and 1.99 (95% CI, 0.52-7.65) for major bleeding. We found no statistically significant heterogeneity among individual study findings. CONCLUSIONS: Our findings are consistent with a 33% reduction in risk of MI or death in patients with unstable angina treated with aspirin plus heparin compared with those treated with aspirin alone. The bulk of evidence suggests that most patients with unstable angina should be treated with both heparin and aspirin.

Question: Therapy

Study Type: Systematic Review

Learner Level: Beginner / Intermediate

Notes: Important paper that directed clinical practice with respect to the use of aspirin and heparin. Good figures for discussing systematic review, confidence intervals, heterogeneity.

Pagliari, L., et al. (1992). "Prevention of first bleeding in cirrhosis. A meta-analysis of randomized trials of nonsurgical treatment." Ann Intern Med 117(1): 59-70.

OBJECTIVE: To assess the effectiveness of beta-blockers and endoscopic sclerotherapy in the prevention of first bleeding and reduction of mortality in patients with cirrhosis and esophagogastric varices. DATA SOURCES: Pertinent studies were selected using MEDLINE (1980 to 1990), reference lists from published articles or reviews, and congress abstract lists. STUDY SELECTION: Randomized trials comparing beta-blockers or sclerotherapy with a nonactive treatment. Nine randomized clinical trials of beta-blockers and 19 trials of sclerotherapy were reviewed. Seven trials of beta-blockers and 15 of sclerotherapy were published as full papers. DATA EXTRACTION: Crude rates of bleeding and death in treated and control groups were extracted from each trial by three independent observers according to the intention-to-treat principle. The quality of published papers was systematically assessed and scored. DATA SYNTHESIS: The Mantel-Haenszel-Peto method was used for statistical evaluation of heterogeneity and for pooling of the results. No substantial heterogeneity was found, and the incidence of bleeding in trials of beta-blockers was significantly reduced (pooled odds ratio, 0.54; 95% CI, 0.39 to 0.74), particularly in patients with large or medium-sized varices or in those with varices and a hepatic vein pressure gradient above 12 mm Hg; however, only a trend toward reduced mortality was obtained. Sclerotherapy trials were highly heterogeneous in the direction of the treatment effects on both bleeding (pooled odds ratio, 0.6; CI, 0.49 to 0.74) and mortality (pooled odds ratio, 0.76; CI, 0.61 to 0.94). The quality of the trials and the rate of bleeding in the untreated groups were the major sources of heterogeneity. The favorable results of sclerotherapy were obtained in trials with high bleeding rates among controls; several of these trials had a low quality score. CONCLUSIONS: Beta-blockers may be recommended for prevention of first bleeding in cirrhotic patients with varices who have a high risk for bleeding. The effectiveness of sclerotherapy remains undetermined. Further trials in high-risk patients may prove useful if improved criteria to predict bleeding risk become available.

Question: Therapy

Study Type: Systematic Review

Learner Level: Intermediate / Advanced

Notes: Very good article for discussion of heterogeneity. (There is significant heterogeneity here...There are many good figures for illustration. However, the many figures and different kind of graphical representations may intimidate those who are not familiar with how to look at a systematic review. This is good for those who wish to take on a more challenging paper, but should be avoided by more novice meta-analysis learners.

Philip, P., et al. (2006). "The effects of coffee and napping on nighttime highway driving: a randomized trial." Ann Intern Med 144(11): 785-791.

BACKGROUND: Sleep-related accidents often involve healthy young persons who are driving at night. Coffee and napping restore alertness, but no study has compared their effects on real nighttime driving performances. OBJECTIVE: To test the effects of 125 mL of coffee (half a cup) containing 200 mg of caffeine, placebo (decaffeinated coffee containing 15 mg of caffeine), or a 30-minute nap (at 1:00 a.m.) in a car on nighttime driving performance. DESIGN: Double-blind, randomized, crossover study. SETTING: Sleep laboratory and open highway. PARTICIPANTS: 12 young men (mean age, 21.3 years [SD, 1.8]). MEASUREMENTS: Self-rated fatigue and sleepiness, inappropriate line crossings from video recordings during highway driving, and polysomnographic recordings during the nap and subsequent sleep. INTERVENTION: Participants drove 200 km (125 miles) between 6:00 p.m. and 7:30 p.m. (daytime reference condition) or between 2:00 a.m. and 3:30 a.m. (coffee, decaffeinated coffee, or nap condition). After intervention, participants returned to the laboratory to sleep. RESULTS: Nighttime driving performance was similar to daytime performance (0 to 1 line crossing) for 75% of participants after coffee (0 or 1 line crossing), for 66% after the nap (P = 0.66 vs. coffee), and for only 13% after placebo (P = 0.041 vs. nap; P = 0.014 vs. coffee). The incidence rate ratios for having a line crossing after placebo were 3.7 (95% CI, 1.2 to 11.0; P = 0.001) compared with coffee and 2.9 (CI, 1.7 to 5.1; P = 0.021) compared with nap. A statistically significant interindividual variability was observed in response to sleep deprivation and countermeasures. Sleep latencies and efficiency during sleep after nighttime driving were similar in the 3 conditions. LIMITATIONS: Only 1 dose of coffee and 1 nap duration were tested. Effects may differ in other patient or age groups. CONCLUSIONS: Drinking coffee or napping at night statistically significantly reduces driving impairment without altering subsequent sleep.

Question: Therapy

Study Type: RCT

Learner Level: Beginner

Notes: This is a randomized, partly blinded cross-over trial in which all drivers performed 4 sessions to allow comparisons. The study looked at effect of coffee and napping strategy on driving behaviors in those who are fatigued or driving at night. Very small sample (12 healthy young men), however for early learners or for illustrating principles of RCT, this trial can be material for great teaching. The topic is relevant to all audiences, especially educators given the 80 hour duty limit. Good paper for teaching results as the comparisons are simple and you can easily calculate NNT for preventing line crossings.

Phung, O. J., et al. (2011). "Dosing frequency of unfractionated heparin thromboprophylaxis: a meta-analysis." *Chest* 140(2): 374-381.

BACKGROUND: In medical patients, it is unclear whether thromboprophylaxis with low-dose unfractionated heparin (UFH) should be administered bid or tid. **METHODS:** This study was a mixed-treatment comparison meta-analysis of randomized control trials that enrolled hospitalized nonsurgical patients at risk for VTE and compared UFH bid, UFH tid, or low-molecular-weight heparin (LMWH) to one another or to an inactive control subject. DVT, pulmonary embolism (PE), major bleeding, and death were measured. A Bayesian framework using a random-effects model was applied. **RESULTS:** Sixteen trials with moderate methodologic quality enrolling 27,667 patients contributed to this analysis. The relative risk and 95% credible intervals comparing UFH tid to UFH bid for DVT, PE, death, and major bleeding were 1.56 (0.64-4.33), 1.67 (0.49-208.09), 1.17 (0.72-1.95), and 0.89 (0.08-7.05), respectively. When compared with either dose of UFH, the use of LMWH has an effect similar to UFH on all four outcomes. **CONCLUSIONS:** Moderate-quality evidence suggests that subcutaneous UFH bid and UFH tid do not differ in effect on DVT, PE, major bleeding, and mortality. Either of the two dosing regimens of UFH or LMWH appears to be a reasonable strategy for thromboprophylaxis in medical patients. A future randomized trial comparing the two doses of UFH is very unlikely, considering the very large sample size that would be required to demonstrate a significant difference, which, if it exists, is undoubtedly small.

Question: Therapy

Study Type: Meta-analysis

Learner Level: Beginner

Notes: .

Raghu, G., et al. (2004). "A placebo-controlled trial of interferon gamma-1b in patients with idiopathic pulmonary fibrosis." *N Engl J Med* 350(2): 125-133.

BACKGROUND: Idiopathic pulmonary fibrosis is a progressive, fatal disease with no known efficacious therapy. **METHODS:** In a double-blind, multinational trial, we randomly assigned 330 patients with idiopathic pulmonary fibrosis that was unresponsive to corticosteroid therapy to receive subcutaneous interferon gamma-1b or placebo. **RESULTS:** Over a median of 58 weeks, interferon gamma-1b therapy did not significantly affect the primary end point of progression-free survival, defined as the time to disease progression or death, and no significant treatment effect was observed on measures of lung function, gas exchange, or the quality of life. Ten percent of patients in the interferon gamma-1b group died, as compared with 17 percent of patients in the placebo group (P=0.08). Treatment with interferon gamma-1b was associated with more frequent constitutional symptoms. However, the rates of treatment adherence and premature discontinuation of treatment were similar in the two groups. More pneumonias were reported among patients in the interferon gamma-1b group, but the incidence of severe or life-threatening respiratory tract infections was similar in the two groups. **CONCLUSIONS:** In a well-defined population of patients with idiopathic pulmonary fibrosis, interferon gamma-1b did not affect progression-free survival, pulmonary function, or the quality of life. Owing to the size and duration of the trial, a clinically significant survival benefit could not be ruled out.

Question: Therapy

Study Type: RCT

Learner Level: Intermediate

Notes: This can be taught as a package. First exercise goes through the user's guide for therapy and the first article. The second exercise goes through the user's guide for meta-analysis. It helps illustrate some of the points of the validity of the systematic review because the learner's have had a sneak peak at the underlying data. Teach with: Part 2: Bajwa EK et al Interferon gamma 1 b therapy in idiopathic pulmonary fibrosis. A meta-analysis NEJM 2004: 350: 125-133.

Raina, P., et al. (1999). "Influence of companion animals on the physical and psychological health of older people: an analysis of a one-year longitudinal study." *J Am Geriatr Soc* 47(3): 323-329.

OBJECTIVE: To examine whether companion animals or attachment to a companion animal was associated with changes in physical and psychological health in older people and whether the relationships between physical and psychological health and human social networks were modified by the presence or absence of a companion animal. **DESIGN:** A 1-year longitudinal study with standardized telephone interview data collected at baseline and repeated at 1-year **SETTING:** Wellington County, Ontario, Canada **PARTICIPANTS:**

An age- and sex stratified random sample (baseline n = 1054; follow-up n = 995) of noninstitutionalized adults aged 65 and older (mean age = 73, SD +/- 6.3) MEASUREMENTS: Social Network Activity was measured using a family and non-family social support scale, participation in an organized social group, involvement in the affairs of the social group, the practice of confiding in others, feelings of loneliness, and the perceived presence of support in a crisis situation. Chronic conditions were measured as the current number of selected health problems. Pet ownership was assessed by the report of owning a dog or a cat and the Lexington Attachment to Pets Scale score. Physical health was assessed as the ability to perform Activities of Daily Living (ADLs). Psychological health was measured as a summed score comprising the level of satisfaction regarding one's health, family and friend relationships, job, finances, life in general, overall happiness, and perceived mental health. Sociodemographic variables assessed include subject age, sex, marital status, living arrangements, education, household income, and major life events. RESULTS: Pet owners were younger, currently married or living with someone, and more physically active than non-pet owners. The ADL level of respondents who did not currently own pets deteriorated more on average (beta = -.270, P = .040) than that of respondents who currently owned pets after adjusting for other variables during the 1-year period. No statistically significant direct association was observed between pet ownership and change in psychological well-being (P > .100). However, pet ownership significantly modified the relationship between social support and the change in psychological well-being (P = .001) over a 1-year period. CONCLUSIONS: The results demonstrate the benefits of pet ownership in maintaining or slightly enhancing ADL levels of older people. However, a more complex relationship was observed between pet ownership and an older person's well-being.

Question: Therapy

Study Type: Cohort

Learner Level: Beginner

Notes: Easy concept for any audience to grasp, regardless of background or comfort level in medicine, so you can focus on methodology. Good discussion points. What practical/logistical problems might make this a difficult question to answer with a RCT? Discuss cohort selection. Discuss alternative ways that a cohort of community-dwelling elderly could be selected and enrolled...and how different cohort selection strategies might yield different results. How does the fact that this is an observational rather than randomized study impact whether we trust the conclusions? How might confounders play a role in the results?

Ray, K. K., et al. (2010). "Statins and all-cause mortality in high-risk primary prevention: a meta-analysis of 11 randomized controlled trials involving 65,229 participants." *Arch Intern Med* 170(12): 1024-1031.

BACKGROUND: Statins have been shown to reduce the risk of all-cause mortality among individuals with clinical history of coronary heart disease. However, it remains uncertain whether statins have similar mortality benefit in a high-risk primary prevention setting. Notably, all systematic reviews to date included trials that in part incorporated participants with prior cardiovascular disease (CVD) at baseline. Our objective was to reliably determine if statin therapy reduces all-cause mortality among intermediate to high-risk individuals without a history of CVD. DATA SOURCES: Trials were identified through computerized literature searches of MEDLINE and Cochrane databases (January 1970-May 2009) using terms related to statins, clinical trials, and cardiovascular end points and through bibliographies of retrieved studies. STUDY SELECTION: Prospective, randomized controlled trials of statin therapy performed in individuals free from CVD at baseline and that reported details, or could supply data, on all-cause mortality. DATA EXTRACTION: Relevant data including the number of patients randomized, mean duration of follow-up, and the number of incident deaths were obtained from the principal publication or by correspondence with the investigators. DATA SYNTHESIS: Data were combined from 11 studies and effect estimates were pooled using a random-effects model meta-analysis, with heterogeneity assessed with the I(2) statistic. Data were available on 65,229 participants followed for approximately 244,000 person-years, during which 2793 deaths occurred. The use of statins in this high-risk primary prevention setting was not associated with a statistically significant reduction (risk ratio, 0.91; 95% confidence interval, 0.83-1.01) in the risk of all-cause mortality. There was no statistical evidence of heterogeneity among studies (I(2) = 23%; 95% confidence interval, 0%-61% [P = .23]). CONCLUSION: This literature-based meta-analysis did not find evidence for the benefit of statin therapy on all-cause mortality in a high-risk primary prevention set-up.

Question: Therapy

Study Type: Meta-analysis

Learner Level: Beginner

Notes: Well done large meta-analysis of the use of statins for primary prevention in high risk patients without cardiac events. Beginner for a meta-analysis in that it has a manageable amount of data presented, forest plots are straightforward with reporting of I-squared

Richeldi, L., et al. (2014). "Efficacy and safety of nintedanib in idiopathic pulmonary fibrosis." *N Engl J Med* 370(22): 2071-2082.

BACKGROUND: Nintedanib (formerly known as BIBF 1120) is an intracellular inhibitor that targets multiple tyrosine kinases. A phase 2 trial suggested that treatment with 150 mg of nintedanib twice daily reduced lung-function decline and acute exacerbations in patients with idiopathic pulmonary fibrosis. METHODS: We conducted two replicate 52-week, randomized, double-blind, phase 3 trials (INPULSIS-1 and INPULSIS-2) to evaluate the efficacy and safety of 150 mg of nintedanib twice daily as compared with placebo in patients with idiopathic pulmonary fibrosis. The primary end point was the annual rate of decline in forced vital capacity (FVC). Key secondary end points were the time to the first acute exacerbation and the change from baseline in the total score on the St. George's Respiratory Questionnaire, both assessed over a 52-week period. RESULTS: A total of 1066 patients were randomly assigned in a 3:2 ratio to receive nintedanib or placebo. The adjusted annual rate of change in FVC was -114.7 ml with nintedanib versus -239.9 ml with placebo (difference, 125.3 ml; 95% confidence interval [CI], 77.7 to 172.8; P<0.001) in INPULSIS-1 and -113.6

ml with nintedanib versus -207.3 ml with placebo (difference, 93.7 ml; 95% CI, 44.8 to 142.7; $P < 0.001$) in INPULSIS-2. In INPULSIS-1, there was no significant difference between the nintedanib and placebo groups in the time to the first acute exacerbation (hazard ratio with nintedanib, 1.15; 95% CI, 0.54 to 2.42; $P = 0.67$); in INPULSIS-2, there was a significant benefit with nintedanib versus placebo (hazard ratio, 0.38; 95% CI, 0.19 to 0.77; $P = 0.005$). The most frequent adverse event in the nintedanib groups was diarrhea, with rates of 61.5% and 18.6% in the nintedanib and placebo groups, respectively, in INPULSIS-1 and 63.2% and 18.3% in the two groups, respectively, in INPULSIS-2. **CONCLUSIONS:** In patients with idiopathic pulmonary fibrosis, nintedanib reduced the decline in FVC, which is consistent with a slowing of disease progression; nintedanib was frequently associated with diarrhea, which led to discontinuation of the study medication in less than 5% of patients. (Funded by Boehringer Ingelheim; INPULSIS-1 and INPULSIS-2 ClinicalTrials.gov numbers, NCT01335464 and NCT01335477.)

Question: Therapy

Study Type: RCT

Learner Level: Beginner

Notes: RCT with good description of methods, good for teaching about surrogate outcomes (vs. patient important outcomes) and role of industry funding.

Ross, J. R., et al. (2003). "Systematic review of role of bisphosphonates on skeletal morbidity in metastatic cancer." *BMJ* 327(7413): 469.

OBJECTIVE: To review the evidence for the use of bisphosphonates to reduce skeletal morbidity in cancer patients with bone metastases. **DATA SOURCES:** Electronic databases, scanning reference lists, and consultation with experts and pharmaceutical companies. Foreign language papers were included. **STUDY SELECTION:** Included trials were randomised controlled trials of patients with malignant disease and bone metastases who were treated with oral or intravenous bisphosphonate compared with another bisphosphonate, placebo, or standard care. All trials measured at least one outcome of skeletal morbidity. **RESULTS:** 95 articles were identified; 30 studies fulfilled inclusion criteria. In studies that lasted ≥ 6 months, compared with placebo bisphosphonates significantly reduced the odds ratio for fractures (vertebral 0.69, 95% confidence interval 0.57 to 0.84, $P < 0.0001$; non-vertebral 0.65, 0.54 to 0.79, $P < 0.0001$; combined 0.65, 0.55 to 0.78, $P < 0.0001$), radiotherapy (0.67, 0.57 to 0.79, $P < 0.0001$), and hypercalcaemia (0.54, 0.36 to 0.81, $P = 0.003$) but not for orthopaedic surgery (0.70, 0.46 to 1.05, $P = 0.086$) or spinal cord compression (0.71, 0.47 to 1.08, $P = 0.113$). The reduction in orthopaedic surgery was significant in studies that lasted over a year (0.59, 0.39 to 0.88, $P = 0.009$). Use of bisphosphonates significantly increased time to first skeletal related event but did not increase survival. Subanalyses showed that most evidence supports use of intravenous aminobisphosphonates. **CONCLUSIONS:** In people with metastatic bone disease bisphosphonates significantly decrease skeletal morbidity, except for spinal cord compression and increased time to first skeletal related event. Treatment should start when bone metastases are diagnosed and continue until it is no longer clinically relevant.

Question: Therapy

Study Type: Meta-analysis

Learner Level: Intermediate / Advanced

Notes: Strong methodology Meta-analysis with very clear description of methods (Figure 1 nice overview for flow of selected papers with on-line access to even more complete details). Good Discussion Points: Lots of great forest plots! / Both endpoints that use straight RR and those needing conversion to effect size / No data table in paper, but available from a website. Generates great discussion about homogeneity and generalizability (includes lytic cancers, blastic cancers, wide population and age range). Good paper to point out pitfall of trying to get a systematic review to answer more than its focused clinical question (e.g., which bisphosphonate should you try first?)

Rowbotham, M., et al. (1998). "Gabapentin for the treatment of postherpetic neuralgia: a randomized controlled trial." *JAMA* 280(21): 1837-1842.

CONTEXT: Postherpetic neuralgia (PHN) is a syndrome of often intractable neuropathic pain following herpes zoster (shingles) that eludes effective treatment in many patients. **OBJECTIVE:** To determine the efficacy and safety of the anticonvulsant drug gabapentin in reducing PHN pain. **DESIGN:** Multicenter, randomized, double-blind, placebo-controlled, parallel design, 8-week trial conducted from August 1996 through July 1997. **SETTING:** Sixteen US outpatient clinical centers. **PARTICIPANTS:** A total of 229 subjects were randomized. **INTERVENTION:** A 4-week titration period to a maximum dosage of 3600 mg/d of gabapentin or matching placebo. Treatment was maintained for another 4 weeks at the maximum tolerated dose. Concomitant tricyclic antidepressants and/or narcotics were continued if therapy was stabilized prior to study entry and remained constant throughout the study. **MAIN OUTCOME MEASURES:** The primary efficacy measure was change in the average daily pain score based on an 11-point Likert scale (0, no pain; 10, worst possible pain) from baseline week to the final week of therapy. Secondary measures included average daily sleep scores, Short-Form McGill Pain Questionnaire (SF-MPQ), Subject Global Impression of Change and investigator-rated Clinical Global Impression of Change, Short Form-36 (SF-36) Quality of Life Questionnaire, and Profile of Mood States (POMS). Safety measures included the frequency and severity of adverse events. **RESULTS:** One hundred thirteen patients received gabapentin, and 89 (78.8%) completed the study; 116 received placebo, and 95 (81.9%) completed the study. By intent-to-treat analysis, subjects receiving gabapentin had a statistically significant reduction in average daily pain score from 6.3 to 4.2 points compared with a change from 6.5 to 6.0 points in subjects randomized to receive placebo ($P < .001$). Secondary measures of pain as well as changes in pain and sleep interference showed improvement with gabapentin ($P < .001$). Many measures within the SF-36 and POMS also significantly favored gabapentin ($P < \text{or} = .01$). Somnolence, dizziness, ataxia, peripheral edema, and infection were all more frequent in the

gabapentin group, but withdrawals were comparable in the 2 groups (15 [13.3%] in the gabapentin group vs 11 [9.5%] in the placebo group). CONCLUSIONS: Gabapentin is effective in the treatment of pain and sleep interference associated with PHN. Mood and quality of life also improve with gabapentin therapy.

Question: Therapy

Study Type: RCT

Learner Level: Beginner / Intermediate

Notes: This is a pivotal study of widely used agent for neuropathic pain. Good for discussing principles of pain research, a very important clinical issue ("the fifth vital sign"); Good Discussion Points: Excellent for discussion of pros and cons of scales as outcome measures; Good to understand strengths and weaknesses of mean scale scores versus percent responder rates between groups when reporting key trial outcomes (note: can't calculate NNT from scale scores) Can teach concept of minimally detectable differences, clinically meaningful differences; Can discuss difference between intention to treat and 'efficacy-evaluable' analyses (Data collection and statistical analysis page 1838)

Saitz, R., et al. (1994). "Individualized treatment for alcohol withdrawal. A randomized double-blind controlled trial." *JAMA* 272(7): 519-523.

OBJECTIVE--To assess the effect of an individualized treatment regimen on the intensity and duration of medication treatment for alcohol withdrawal. **DESIGN--**A randomized double-blind, controlled trial. **SETTING--**An inpatient detoxification unit in a Veterans Affairs medical center. **PATIENTS--**One hundred one patients admitted for the treatment of alcohol withdrawal who could give informed consent and had no history of seizures or medication use that might alter the clinical course of withdrawal. **INTERVENTION--**Patients were randomized to either a standard course of chlordiazepoxide four times daily with additional medication as needed (fixed-schedule therapy) or to a treatment regimen that provided chlordiazepoxide only in response to the development of the signs and symptoms of alcohol withdrawal (symptom-triggered therapy). The need for administration of "as-needed" medication was determined using a validated measure of the severity of alcohol withdrawal. **MAIN OUTCOME MEASURES--**Duration of medication treatment and total chlordiazepoxide administered. **RESULTS--**The median duration of treatment in the symptom-triggered group was 9 hours, compared with 68 hours in the fixed-schedule group ($P < .001$). The symptom-triggered group received 100 mg of chlordiazepoxide, and the fixed-schedule group received 425 mg ($P < .001$). There were no significant differences in the severity of withdrawal during treatment or in the incidence of seizures or delirium tremens. **CONCLUSIONS--**Symptom-triggered therapy individualizes treatment, decreases both treatment duration and the amount of benzodiazepine used, and is as efficacious as standard fixed-schedule therapy for alcohol withdrawal.

Question: Therapy

Study Type: RCT

Learner Level: Beginner

Notes: Good to teach concealed allocation

Salpeter, S. R., et al. (2004). "Cardiovascular effects of beta-agonists in patients with asthma and COPD: a meta-analysis." *Chest* 125(6): 2309-2321.

BACKGROUND: beta-Adrenergic agonists exert physiologic effects that are the opposite of those of beta-blockers. beta-Blockers are known to reduce morbidity and mortality in patients with cardiac disease. beta(2)-Agonist use in patients with obstructive airway disease has been associated with an increased risk for myocardial infarction, congestive heart failure, cardiac arrest, and acute cardiac death. **OBJECTIVES:** To assess the cardiovascular safety of beta(2)-agonist use in patients with obstructive airway disease, defined as asthma or COPD. **METHODS:** A meta-analysis of randomized placebo-controlled trials of beta(2)-agonist treatment in patients with obstructive airway disease was performed, to evaluate the short-term effect on heart rate and potassium concentrations, and the long-term effect on adverse cardiovascular events. Longer duration trials were included in the analysis if they reported at least one adverse event. Adverse events included sinus and ventricular tachycardia, syncope, atrial fibrillation, congestive heart failure, myocardial infarction, cardiac arrest, or sudden death. **RESULTS:** Thirteen single-dose trials and 20 longer duration trials were included in the study. A single dose of beta(2)-agonist increased the heart rate by 9.12 beats/min (95% confidence interval [CI], 5.32 to 12.92) and reduced the potassium concentration by 0.36 mmol/L (95% CI, 0.18 to 0.54), compared to placebo. For trials lasting from 3 days to 1 year, beta(2)-agonist treatment significantly increased the risk for a cardiovascular event (relative risk [RR], 2.54; 95% CI, 1.59 to 4.05) compared to placebo. The RR for sinus tachycardia alone was 3.06 (95% CI, 1.70 to 5.50), and for all other events it was 1.66 (95% CI, 0.76 to 3.6). **CONCLUSION:** beta(2)-Agonist use in patients with obstructive airway disease increases the risk for adverse cardiovascular events. The initiation of treatment increases heart rate and reduces potassium concentrations compared to placebo. It could be through these mechanisms, and other effects of beta-adrenergic stimulation, that beta(2)-agonists may precipitate ischemia, congestive heart failure, arrhythmias, and sudden death.

Question: Therapy

Study Type: Meta-analysis

Learner Level: Beginner

Notes: Good to teach surrogate outcomes

Salpeter, S. R., et al. (2010). "Long-acting beta-agonists with and without inhaled corticosteroids and catastrophic asthma events." *Am J Med* 123(4): 322-328 e322.

BACKGROUND: It is unclear whether long-acting beta-agonists with concomitant inhaled corticosteroids increase asthma-related intubations and deaths. We pooled data on long-acting beta-agonists with variable and concomitant inhaled corticosteroids to evaluate the risk for catastrophic asthma events. **METHODS:** We conducted searches of electronic databases, the US Food and Drug Administration website, clinical-trials registries, and selected references through December 2008. We analyzed randomized controlled trials in patients with asthma, which lasted at least 3 months, evaluated long-acting beta-agonists compared with placebo or long-acting beta-agonists with inhaled corticosteroids compared with corticosteroids alone, and included at least 1 catastrophic event, defined as asthma-related intubation or death. **RESULTS:** In pooled trial data that included 36,588 participants, long-acting beta-agonists increased catastrophic events 2-fold (Peto odds ratio [OR] 2.10; 95% confidence interval [CI], 1.37-3.22). Statistically significant increases were seen for long-acting beta-agonists with variable corticosteroids compared with placebo (OR 1.83; 95% CI, 1.14-2.95) and for concomitant treatment with corticosteroids compared with corticosteroids alone (OR 3.65; 95% CI, 1.39-9.55). Similar increases in risk were seen for variable and concomitant corticosteroid use, salmeterol and formoterol, and children and adults. When the analysis was restricted to trials with controlled corticosteroid use, given as part of the study intervention, concomitant treatment still increased catastrophic events compared with corticosteroids alone (OR 8.19; 95% CI, 1.10-61.18). **CONCLUSION:** Long-acting beta-agonists increase the risk for asthma-related intubations and deaths, even when used in a controlled fashion with concomitant inhaled corticosteroids.

Question: Therapy

Study Type: Meta-analysis

Learner Level: Beginner - Intermediate

Notes: controversial topic; RCTs are summarized with outcome of increased risk (harm as opposed to the usual framework of looking for benefit); Good to discuss ethics (the RCTs were done to show benefit, but didn't end up showing that in all cases...); Nice Forest plots including use of I-squared; Article accompanied by an ACP-JC

Santangeli, P., et al. (2010). "Meta-analysis: age and effectiveness of prophylactic implantable cardioverter-defibrillators." *Ann Intern Med* 153(9): 592-599.

BACKGROUND: Implantable cardioverter-defibrillators (ICDs) for the primary prevention of sudden cardiac death have been proven effective in several clinical trials. **PURPOSE:** To summarize evidence about the effectiveness of ICDs versus standard medical therapy for the primary prevention of sudden cardiac death in different age groups of patients with severe left ventricular dysfunction. **DATA SOURCES:** MEDLINE, Embase, CENTRAL, BioMed Central, Cardiosource, ClinicalTrials.gov, and ISI Web of Science (January 1970 to April 2010) were searched with no language restrictions. **STUDY SELECTION:** Two independent reviewers screened titles and abstracts to identify randomized, controlled trials of prophylactic ICD versus medical therapy in patients with severe left ventricular dysfunction that provided data about mortality outcomes for different age groups. **DATA EXTRACTION:** Two independent reviewers assessed risk for bias of trials and extracted patient and study characteristics and hazard ratios (HRs) relevant to all-cause mortality. **DATA SYNTHESIS:** Five trials (MADIT-II, DEFINITE, DINAMIT, SCD-HeFT, and IRIS) that enrolled 5783 patients (44% were elderly) were included. The primary analysis, which excluded the 2 trials enrolling patients early after acute myocardial infarction (DINAMIT and IRIS), found that prophylactic ICD therapy reduced mortality in younger patients (HR, 0.65 [95% CI, 0.50 to 0.83]; $P < 0.001$). A smaller survival benefit was found in elderly patients (HR, 0.75 [95% CI, 0.61 to 0.91]) that was not confirmed when MADIT-II patients older than 70 years were excluded or when data from DINAMIT and IRIS were included [corrected]. **LIMITATIONS:** Four potentially eligible trials were not included in the meta-analysis because mortality data by age group were not available. Adjustment for differences in comorbid conditions and medical therapies among patients enrolled in the trials was not possible. **CONCLUSION:** Available data suggest that prophylactic ICD therapy may be less beneficial for elderly patients with severe left ventricular dysfunction than for younger patients [corrected]. **PRIMARY FUNDING SOURCE:** None.

Question: Therapy

Study Type: Meta-analysis

Learner Level: Intermediate

Notes: Nice example of a systematic review in examining subgroup efficacy. Good clear assessment of evidence quality, forest plots, discussion points on heterogeneity. Moderate difficulty.

Self, W. H., et al. (2018). "Balanced Crystalloids versus Saline in Noncritically Ill Adults." *N Engl J Med* 378(9): 819-828.

BACKGROUND: Comparative clinical effects of balanced crystalloids and saline are uncertain, particularly in noncritically ill patients cared for outside an intensive care unit (ICU). **METHODS:** We conducted a single-center, pragmatic, multiple-crossover trial comparing balanced crystalloids (lactated Ringer's solution or Plasma-Lyte A) with saline among adults who were treated with intravenous crystalloids in the emergency department and were subsequently hospitalized outside an ICU. The type of crystalloid that was administered in the emergency department was assigned to each patient on the basis of calendar month, with the entire emergency department crossing over between balanced crystalloids and saline monthly during the 16-month trial. The primary outcome was hospital-free days (days alive after discharge before day 28). Secondary outcomes included major adverse kidney events within 30 days - a composite of death from any cause, new renal-replacement therapy, or persistent renal dysfunction

(defined as an elevation of the creatinine level to $\geq 200\%$ of baseline) - all censored at hospital discharge or 30 days, whichever occurred first. RESULTS: A total of 13,347 patients were enrolled, with a median crystalloid volume administered in the emergency department of 1079 ml and 88.3% of the patients exclusively receiving the assigned crystalloid. The number of hospital-free days did not differ between the balanced-crystalloids and saline groups (median, 25 days in each group; adjusted odds ratio with balanced crystalloids, 0.98; 95% confidence interval [CI], 0.92 to 1.04; $P=0.41$). Balanced crystalloids resulted in a lower incidence of major adverse kidney events within 30 days than saline (4.7% vs. 5.6%; adjusted odds ratio, 0.82; 95% CI, 0.70 to 0.95; $P=0.01$). CONCLUSIONS: Among noncritically ill adults treated with intravenous fluids in the emergency department, there was no difference in hospital-free days between treatment with balanced crystalloids and treatment with saline. (Funded by the Vanderbilt Institute for Clinical and Translational Research and others; SALT-ED ClinicalTrials.gov number, NCT02614040 .).

Question: RCT

Study Type: Therapy

Learner Level: Beginner

Notes: Good for teaching study design (e.g., are you happy with how it was randomized? Were enough patients enrolled? Etc.).

Shekelle, P. G., et al. (2004). "Effect of supplemental vitamin E for the prevention and treatment of cardiovascular disease." *Journal of General Internal Medicine* 19(4): 380-389.

OBJECTIVE: To evaluate and synthesize the evidence on the effect of supplements of vitamin E on the prevention and treatment of cardiovascular disease. **DESIGN:** Systematic review of placebo-controlled randomized controlled trials; meta-analysis where justified. **MEASUREMENTS AND MAIN RESULTS:** Eighty-four eligible trials were identified. For the outcomes of all-cause mortality, cardiovascular mortality, fatal or nonfatal myocardial infarction, and blood lipids, neither supplements of vitamin E alone nor vitamin E given with other agents yielded a statistically significant beneficial or adverse pooled relative risk (for example, pooled relative risk of vitamin E alone = 0.96 [95% confidence interval (CI), 0.84 to 1.10]; 0.97 [95% CI, 0.80 to 1.90]; and 0.72 [95% CI, 0.51 to 1.02] for all-cause mortality, cardiovascular mortality, and nonfatal myocardial infarction, respectively. **CONCLUSIONS:** There is good evidence that vitamin E supplementation does not beneficially or adversely affect cardiovascular outcomes.

Question: Therapy

Study Type: Systematic Review

Learner Level: Beginner

Notes: High quality systematic review. Easy to understand therapy and outcomes if teaching a group from mixed specialties. It is a treatment that many people are still on contrary to the evidence. A little bit complicated since it looked at vitamin E alone and vitamin E in combination as 2 separate analyses in the same paper. Good Discussion Points: Validity of SR since it is extremely well done with extensive search and description of methodology. Discussion of heterogeneity. Main outcome did not show heterogeneity, but the secondary outcomes did. This can be seen statistically as well as in the Forest plots. They also mention the I² statistic as well as χ^2 , so can use this for advanced learners. Given the differences between studies it is a good example of when to pool and when not to. I focus on the main outcome of mortality to try and simplify this long article. ACP Journal Club reviews this article. See also Miller, E. *Annals of Internal Med* Jan 4 2005 and Eidelman RS. *Annals of Internal Med* 2004.

Shelton, R. C., et al. (2001). "Effectiveness of St John's wort in major depression: a randomized controlled trial." *JAMA* 285(15): 1978-1986.

CONTEXT: Extracts of St John's wort are widely used to treat depression. Although more than 2 dozen clinical trials have been conducted with St John's wort, most have significant flaws in design and do not enable meaningful interpretation. **OBJECTIVE:** To compare the efficacy and safety of a standardized extract of St John's wort with placebo in outpatients with major depression. **DESIGN AND SETTING:** Randomized, double-blind, placebo-controlled clinical trial conducted between November 1998 and January 2000 in 11 academic medical centers in the United States. **PARTICIPANTS:** Two hundred adult outpatients (mean age, 42.4 years; 67.0% female; 85.9% white) diagnosed as having major depression and having a baseline Hamilton Rating Scale for Depression (HAM-D) score of at least 20. **INTERVENTION:** Participants completed a 1-week, single-blind run-in of placebo, then were randomly assigned to receive either St John's wort extract ($n = 98$; 900 mg/d for 4 weeks, increased to 1200 mg/d in the absence of an adequate response thereafter) or placebo ($n = 102$) for 8 weeks. **MAIN OUTCOME MEASURES:** The primary outcome measure was rate of change on the HAM-D over the treatment period. Secondary measures included the Beck Depression Inventory (BDI), Hamilton Rating Scale for Anxiety (HAM-A), the Global Assessment of Function (GAF) scale, and the Clinical Global Impression-Severity and -Improvement scales (CGI-S and CGI-I). **RESULTS:** The random coefficient analyses for the HAM-D, HAM-A, CGI-S, and CGI-I all showed significant effects for time but not for treatment or time-by-treatment interaction (for HAM-D scores, $P<.001$, $P = .16$, and $P = .58$, respectively). Analysis of covariance showed nonsignificant effects for BDI and GAF scores. The proportion of participants achieving an a priori definition of response did not differ between groups. The number reaching remission of illness was significantly higher with St John's wort than with placebo ($P = .02$), but the rates were very low in the full intention-to-treat analysis (14/98 [14.3%] vs 5/102 [4.9%], respectively). St John's wort was safe and well tolerated. Headache was the only adverse event that occurred with greater frequency with St John's wort than placebo (39/95 [41%] vs 25/100 [25%], respectively). **CONCLUSION:** In this study, St John's wort was not effective for treatment of major depression.

Question: Therapy

Study Type: RCT

Learner Level: Beginner / Intermediate

Notes: Strong Methodology, very well described Interesting, topic (herbal preparations) that have broad impact due to OTC availability; Largely an equivalence trial, however there are some differences so you can calculate NNTs for those measures. Good discussion points: Nice Trial Flow design (fig 1) for discussion on methods; Can discuss intention to treat Has a great table of prior evidence with study limitations listed; GREAT for discussion of methodology issues. Prior reports have shown impact of St. John's Wort, so it is interesting to discuss why this paper might have different findings (i.e., different methods? Different populations of patients? Greater numbers of patients?)

Shepherd, J., et al. (2002). "Pravastatin in elderly individuals at risk of vascular disease (PROSPER): a randomised controlled trial." *Lancet* 360(9346): 1623-1630.

BACKGROUND: Although statins reduce coronary and cerebrovascular morbidity and mortality in middle-aged individuals, their efficacy and safety in elderly people is not fully established. Our aim was to test the benefits of pravastatin treatment in an elderly cohort of men and women with, or at high risk of developing, cardiovascular disease and stroke. **METHODS:** We did a randomised controlled trial in which we assigned 5804 men (n=2804) and women (n=3000) aged 70-82 years with a history of, or risk factors for, vascular disease to pravastatin (40 mg per day; n=2891) or placebo (n=2913). Baseline cholesterol concentrations ranged from 4.0 mmol/L to 9.0 mmol/L. Follow-up was 3.2 years on average and our primary endpoint was a composite of coronary death, non-fatal myocardial infarction, and fatal or non-fatal stroke. Analysis was by intention-to-treat. **FINDINGS:** Pravastatin lowered LDL cholesterol concentrations by 34% and reduced the incidence of the primary endpoint to 408 events compared with 473 on placebo (hazard ratio 0.85, 95% CI 0.74-0.97, p=0.014). Coronary heart disease death and non-fatal myocardial infarction risk was also reduced (0.81, 0.69-0.94, p=0.006). Stroke risk was unaffected (1.03, 0.81-1.31, p=0.8), but the hazard ratio for transient ischaemic attack was 0.75 (0.55-1.00, p=0.051). New cancer diagnoses were more frequent on pravastatin than on placebo (1.25, 1.04-1.51, p=0.020). However, incorporation of this finding in a meta-analysis of all pravastatin and all statin trials showed no overall increase in risk. Mortality from coronary disease fell by 24% (p=0.043) in the pravastatin group. Pravastatin had no significant effect on cognitive function or disability. **INTERPRETATION:** Pravastatin given for 3 years reduced the risk of coronary disease in elderly individuals. PROSPER therefore extends to elderly individuals the treatment strategy currently used in middle aged people.

Question: Therapy

Study Type: RCT

Learner Level: Beginner / Intermediate

Notes: Excellent randomized controlled trial methodology applicable to the elderly. Has an ACP journal club summary of the paper and results, which can be used for teaching. Nice Hazard plots are a good way to demonstrate effect size and confidence intervals. Good discussion points: Fantastic discussion paper for the difference between validity and applicability. Outstanding methods that show a treatment effect in this older population. However the decision making process for whether one would want to treat someone who was 80 years old and had risk factors but no known CAD is quite another story... Fun teaching exercise is to have several different case-patients with different characteristics and values. Have the trainees apply the same evidence to different patients. Can discuss the balance of benefits of therapy vs. potential harms and costs both to individual patients and to society.

Simon, J. A., et al. (2003). "Intensive smoking cessation counseling versus minimal counseling among hospitalized smokers treated with transdermal nicotine replacement: a randomized trial." *Am J Med* 114(7): 555-562.

PURPOSE: To determine whether an intensive cognitive-behavioral intervention begun during hospitalization when combined with transdermal nicotine replacement therapy is more effective than a minimal counseling intervention combined with transdermal nicotine replacement therapy in helping inpatients to quit smoking. **METHODS:** A total of 223 patients who smoked were enrolled in a hospital-based randomized smoking cessation trial at the San Francisco Veterans Affairs Medical Center. One hundred and seven participants (48%) received intensive counseling and outpatient telephone follow-up; 116 participants (52%) received minimal counseling. All study participants received 2 months of transdermal nicotine replacement therapy. We determined 6-month quit rates by self-report and measured saliva cotinine levels or obtained proxy reports to confirm self-reported smoking cessation at 12 months. Analyses adjusted for baseline differences in the distribution of coronary disease. **RESULTS:** At 6 months, 35% (36/103) of the intensive intervention group reported quitting, compared with 21% (23/109) of the comparison group (relative risk [RR] = 1.7; 95% confidence interval [CI]: 1.1 to 2.7). At 12 months, the self-reported quit rate was 33% (33/99) in the intensive intervention group versus 20% (21/103) in the comparison group (RR = 1.7; 95% CI: 1.1 to 2.7). Based on biochemical or proxy confirmation, 29% (30/102) in the intensive intervention group versus 20% (21/107) in the comparison group quit smoking at 12 months (RR = 1.6; 95% CI: 0.96 to 2.5). **CONCLUSION:** Hospital-initiated smoking cessation interventions that include transdermal nicotine replacement therapy can improve long-term quit rates.

Question: Therapy

Study Type: RCT

Learner Level: Beginner/Intermediate

Notes: Therapy

Sirven, J. I., et al. (2004). "Seizure prophylaxis in patients with brain tumors: a meta-analysis." *Mayo Clin Proc* 79(12): 1489-1494.

OBJECTIVE: To assess whether antiepileptic drugs (AEDs) should be prescribed to patients with brain tumors who have no history of seizures. **METHODS:** We performed a meta-analysis of randomized controlled trials (1966-2004) that evaluated the efficacy of AED prophylaxis vs no treatment or placebo to prevent seizures in patients with brain tumors who had no history of epilepsy. Summary odds ratios (ORs) were calculated using a random-effects model. Three subanalyses were performed to assess pooled ORs of seizures in patients with primary glial tumors, cerebral metastases, and meningiomas. **RESULTS:** Of 474 articles found in the initial search, 17 were identified as primary studies. Five trials met inclusion criteria: patients with a neoplasm (primary glial tumors, cerebral metastases, and meningiomas) but no history of epilepsy who were randomized to either an AED or placebo. The 3 AEDs studied were phenobarbital, phenytoin, and valproic acid. Of the 5 trials, 4 showed no statistical benefit of seizure prophylaxis with an AED. Meta-analysis confirmed the lack of AED benefit at 1 week (OR, 0.91; 95% confidence interval [CI], 0.45-1.83) and at 6 months (OR, 1.01; 95% CI, 0.51-1.98) of follow-up. The AEDs had no effect on seizure prevention for specific tumor pathology, including primary glial tumors (OR, 3.46; 95% CI, 0.32-37.47), cerebral metastases (OR, 2.50; 95% CI, 0.25-24.72), and meningiomas (OR, 0.62; 95% CI, 0.10-3.85). **CONCLUSIONS:** No evidence supports AED prophylaxis with phenobarbital, phenytoin, or valproic acid in patients with brain tumors and no history of seizures, regardless of neoplastic type. Subspecialists who treat patients with brain tumors need more education on this issue. Future randomized controlled trials should address whether any of the newer AEDs are useful for seizure prophylaxis.

Question: Therapy

Study Type: Meta-analysis

Learner Level: Intermediate

Notes: Good to teach creating your own forest plot

Spiro, D. M., et al. (2006). "Wait-and-see prescription for the treatment of acute otitis media: a randomized controlled trial." *JAMA* 296(10): 1235-1241.

CONTEXT: Acute otitis media (AOM) is the most common diagnosis for which antibiotics are prescribed for children. Previous trials that have evaluated a "wait-and-see prescription" (WASP) for antibiotics, with which parents are asked not to fill the prescription unless the child either is not better or is worse in 48 hours, have excluded children with severe AOM. None of these trials were conducted in an emergency department. **OBJECTIVES:** To determine whether treatment of AOM using a WASP significantly reduces use of antibiotics compared with a "standard prescription" (SP) and to evaluate the effects of this intervention on clinical symptoms and adverse outcomes related to antibiotic use. **DESIGN, SETTING, AND PATIENTS:** A randomized controlled trial conducted between July 12, 2004, and July 11, 2005. Children with AOM aged 6 months to 12 years seen in an emergency department were randomly assigned to receive either a WASP or an SP. All patients received ibuprofen and otic analgesic drops for use at home. A research assistant, blinded to group assignment, conducted structured phone interviews 4 to 6, 11 to 14, and 30 to 40 days after enrollment to determine outcomes. **MAIN OUTCOME MEASURES:** Filling of the antibiotic prescription and clinical course. **RESULTS:** Overall, 283 patients were randomized either to the WASP group (n = 138) or the SP group (n = 145). Substantially more parents in the WASP group did not fill the antibiotic prescription (62% vs 13%; P<.001). There was no statistically significant difference between the groups in the frequency of subsequent fever, otalgia, or unscheduled visits for medical care. Within the WASP group, both fever (relative risk [RR], 2.95; 95% confidence interval [CI], 1.75 - 4.99; P<.001) and otalgia (RR, 1.62; 95% CI, 1.26 - 2.03; P<.001) were associated with filling the prescription. **CONCLUSION:** The WASP approach substantially reduced unnecessary use of antibiotics in children with AOM seen in an emergency department and may be an alternative to routine use of antimicrobials for treatment of such children. **TRIAL REGISTRATION:** clinicaltrials.gov Identifier: NCT00250900.

Question: Therapy

Study Type: RCT

Learner Level: Beginner/intermediate

Notes: Relevant topic for pediatricians or acute care / ER providers (or anyone who has ever had a child with otitis media!) Clear simple methods with impressive results and easy math (NNT = approximately 2 for not filling antibiotic prescriptions). Paper also has very descriptive methods that allow discussion of the randomization process as well as blinding. Clear reporting as well as flow chart as well as description of intention to treat and "worst case sensitivity analysis"

Stevens, B., et al. (1997). "The efficacy of sucrose for relieving procedural pain in neonates--a systematic review and meta-analysis." *Acta Paediatrica* 86(8): 837-842.

The objective was to determine the efficacy and optimal dose of sucrose for relieving procedural pain in neonates. Data were obtained using MEDLINE, EMBASE, Reference Update and personal files and assessed for quality of the methods. Data from all randomized controlled trials where term and preterm neonates received a heelstick or venipuncture were examined for the efficacy of different sucrose doses (0.18 g, 0.24 g, 0.48 g or 0.50 g, 1.0 g) and water (placebo). The primary outcome was the proportion of time crying during 3 min after the painful stimulus. Data were combined across studies using a random effects model, adapted for use with single groups, producing a point estimate and 95% confidence interval (CI). Thirteen trials were identified; eight were rejected as data were inappropriate, non-extractable, or the primary outcome was not measured. Five studies provided data on 271 infants. The proportion of time crying did not differ between 0.18 g of sucrose and water (p > 0.05) but was significantly lower in all other sucrose groups. There were no differences in proportion of time crying between term and preterm neonates. Sucrose reduced

the proportion of time crying during painful procedures in neonates. The 0.18 g dose of sucrose was ineffective. Doses of 0.24 g (2 ml of 12% sucrose solution) were most effective. A dose of 0.50 g provided no additional benefit.

Question: Therapy

Study Type: Systematic Review

Learner Level: Beginner / Intermediate

Notes: Straightforward methods that describe both the selection of articles as well as the assessment of quality of the methods. Good discussion points: Good paper for anyone who has ever had a child that underwent a procedure.

Sung, J. J., et al. (2010). "Continuation of low-dose aspirin therapy in peptic ulcer bleeding: a randomized trial." *Ann Intern Med* 152(1): 1-9.

BACKGROUND: It is uncertain whether aspirin therapy should be continued after endoscopic hemostatic therapy in patients who develop peptic ulcer bleeding while receiving low-dose aspirin. **OBJECTIVE:** To test that continuing aspirin therapy with proton-pump inhibitors after endoscopic control of ulcer bleeding was not inferior to stopping aspirin therapy, in terms of recurrent ulcer bleeding in adults with cardiovascular or cerebrovascular diseases. **DESIGN:** A parallel randomized, placebo-controlled noninferiority trial, in which both patients and clinicians were blinded to treatment assignment, was conducted from 2003 to 2006 by using computer-generated numbers in concealed envelopes. (ClinicalTrials.gov registration number: NCT00153725) **SETTING:** A tertiary endoscopy center. **PATIENTS:** Low-dose aspirin recipients with peptic ulcer bleeding. **INTERVENTION:** 78 patients received aspirin, 80 mg/d, and 78 received placebo for 8 weeks immediately after endoscopic therapy. All patients received a 72-hour infusion of pantoprazole followed by oral pantoprazole. All patients completed follow-up. **MEASUREMENTS:** The primary end point was recurrent ulcer bleeding within 30 days confirmed by endoscopy. Secondary end points were all-cause and specific-cause mortality in 8 weeks. **RESULTS:** 156 patients were included in an intention-to-treat analysis. Three patients withdrew from the trial before finishing follow-up. Recurrent ulcer bleeding within 30 days was 10.3% in the aspirin group and 5.4% in the placebo group (difference, 4.9 percentage points [95% CI, -3.6 to 13.4 percentage points]). Patients who received aspirin had lower all-cause mortality rates than patients who received placebo (1.3% vs. 12.9%; difference, 11.6 percentage points [CI, 3.7 to 19.5 percentage points]). Patients in the aspirin group had lower mortality rates attributable to cardiovascular, cerebrovascular, or gastrointestinal complications than patients in the placebo group (1.3% vs. 10.3%; difference, 9 percentage points [CI, 1.7 to 16.3 percentage points]). **LIMITATIONS:** The sample size is relatively small, and only low-dose aspirin, 80 mg, was used. Two patients with recurrent bleeding in the placebo group did not have further endoscopy. **CONCLUSION:** Among low-dose aspirin recipients who had peptic ulcer bleeding, continuous aspirin therapy may increase the risk for recurrent bleeding but potentially reduces mortality rates. Larger trials are needed to confirm these findings.

Question: Therapy

Study Type: RCT

Learner Level: .

Notes: Very well reported trial on an increasingly relevant clinical question. Excellent example of intention to treat methodology. Can be used to do "therapy math": number needed to treat etc.

Taylor, A. L., et al. (2004). "Combination of isosorbide dinitrate and hydralazine in blacks with heart failure." *N Engl J Med* 351(20): 2049-2057.

BACKGROUND: We examined whether a fixed dose of both isosorbide dinitrate and hydralazine provides additional benefit in blacks with advanced heart failure, a subgroup previously noted to have a favorable response to this therapy. **METHODS:** A total of 1050 black patients who had New York Heart Association class III or IV heart failure with dilated ventricles were randomly assigned to receive a fixed dose of isosorbide dinitrate plus hydralazine or placebo in addition to standard therapy for heart failure. The primary end point was a composite score made up of weighted values for death from any cause, a first hospitalization for heart failure, and change in the quality of life. **RESULTS:** The study was terminated early owing to a significantly higher mortality rate in the placebo group than in the group given isosorbide dinitrate plus hydralazine (10.2 percent vs. 6.2 percent, $P=0.02$). The mean primary composite score was significantly better in the group given isosorbide dinitrate plus hydralazine than in the placebo group ($-0.1+/-1.9$ vs. $-0.5+/-2.0$, $P=0.01$; range of possible values, -6 to +2), as were its individual components (43 percent reduction in the rate of death from any cause [hazard ratio, 0.57; $P=0.01$] 33 percent relative reduction in the rate of first hospitalization for heart failure [16.4 percent vs. 22.4 percent, $P=0.001$], and an improvement in the quality of life [change in score, $-5.6+/-20.6$ vs. $-2.7+/-21.2$, with lower scores indicating better quality of life; $P=0.02$; range of possible values, 0 to 105]). **CONCLUSIONS:** The addition of a fixed dose of isosorbide dinitrate plus hydralazine to standard therapy for heart failure including neurohormonal blockers is efficacious and increases survival among black patients with advanced heart failure.

Question: Therapy

Study Type: RCT

Learner Level: Beginner / Intermediate

Notes: Also referred to as the A-HEFT Trial (African American Heart Failure Trial), this is an RCT of isosorbide / hydralazine in participants who self-identified as black (defined as of African Descent). ACP Journal Club Summary 2005; 142 (2): 37. Good

discussion points: Controversial motivation: NitroMed, the company that holds the patent on the fixed dose combination drug, sponsored the trial. This raises some ethical questions about the motivation for performing the trial in a single racial group as a mechanism to obtain more rapid approval from the FDA (see NEJM Perspective that accompanied the article when it was published. NEJM 2004; 351:2035-37) Controversial recommendations: Should one make race-based recommendations, given proven efficacy in self-identified Blacks and lack of published efficacy in Caucasians? Straightforward Therapy calculations: Can easily calculate NNT for All-cause mortality and for first hospitalization for CHF. Can discuss the concept of Composite Score. Why would the authors have planned the analysis that way? This might lead to a general discussion of the pros and cons of combined outcomes.

Temel, J. S., et al. (2010). "Early palliative care for patients with metastatic non-small-cell lung cancer." *N Engl J Med* 363(8): 733-742.

BACKGROUND: Patients with metastatic non-small-cell lung cancer have a substantial symptom burden and may receive aggressive care at the end of life. We examined the effect of introducing palliative care early after diagnosis on patient-reported outcomes and end-of-life care among ambulatory patients with newly diagnosed disease. **METHODS:** We randomly assigned patients with newly diagnosed metastatic non-small-cell lung cancer to receive either early palliative care integrated with standard oncologic care or standard oncologic care alone. Quality of life and mood were assessed at baseline and at 12 weeks with the use of the Functional Assessment of Cancer Therapy-Lung (FACT-L) scale and the Hospital Anxiety and Depression Scale, respectively. The primary outcome was the change in the quality of life at 12 weeks. Data on end-of-life care were collected from electronic medical records. **RESULTS:** Of the 151 patients who underwent randomization, 27 died by 12 weeks and 107 (86% of the remaining patients) completed assessments. Patients assigned to early palliative care had a better quality of life than did patients assigned to standard care (mean score on the FACT-L scale [in which scores range from 0 to 136, with higher scores indicating better quality of life], 98.0 vs. 91.5; $P=0.03$). In addition, fewer patients in the palliative care group than in the standard care group had depressive symptoms (16% vs. 38%, $P=0.01$). Despite the fact that fewer patients in the early palliative care group than in the standard care group received aggressive end-of-life care (33% vs. 54%, $P=0.05$), median survival was longer among patients receiving early palliative care (11.6 months vs. 8.9 months, $P=0.02$). **CONCLUSIONS:** Among patients with metastatic non-small-cell lung cancer, early palliative care led to significant improvements in both quality of life and mood. As compared with patients receiving standard care, patients receiving early palliative care had less aggressive care at the end of life but longer survival. (Funded by an American Society of Clinical Oncology Career Development Award and philanthropic gifts; ClinicalTrials.gov number, NCT01038271.)

Question: Therapy

Study Type: RCT

Learner Level: Beginner

Notes: Systems intervention RCT of a different care paradigm; Nicely done methods and interesting findings; Can calculate NNT to prevent depression

Thursz, M. R., et al. (2015). "Prednisolone or pentoxifylline for alcoholic hepatitis." *N Engl J Med* 372(17): 1619-1628.

BACKGROUND: Alcoholic hepatitis is a clinical syndrome characterized by jaundice and liver impairment that occurs in patients with a history of heavy and prolonged alcohol use. The short-term mortality among patients with severe disease exceeds 30%. Prednisolone and pentoxifylline are both recommended for the treatment of severe alcoholic hepatitis, but uncertainty about their benefit persists. **METHODS:** We conducted a multicenter, double-blind, randomized trial with a 2-by-2 factorial design to evaluate the effect of treatment with prednisolone or pentoxifylline. The primary end point was mortality at 28 days. Secondary end points included death or liver transplantation at 90 days and at 1 year. Patients with a clinical diagnosis of alcoholic hepatitis and severe disease were randomly assigned to one of four groups: a group that received a pentoxifylline-matched placebo and a prednisolone-matched placebo, a group that received prednisolone and a pentoxifylline-matched placebo, a group that received pentoxifylline and a prednisolone-matched placebo, or a group that received both prednisolone and pentoxifylline. **RESULTS:** A total of 1103 patients underwent randomization, and data from 1053 were available for the primary end-point analysis. Mortality at 28 days was 17% (45 of 269 patients) in the placebo-placebo group, 14% (38 of 266 patients) in the prednisolone-placebo group, 19% (50 of 258 patients) in the pentoxifylline-placebo group, and 13% (35 of 260 patients) in the prednisolone-pentoxifylline group. The odds ratio for 28-day mortality with pentoxifylline was 1.07 (95% confidence interval [CI], 0.77 to 1.49; $P=0.69$), and that with prednisolone was 0.72 (95% CI, 0.52 to 1.01; $P=0.06$). At 90 days and at 1 year, there were no significant between-group differences. Serious infections occurred in 13% of the patients treated with prednisolone versus 7% of those who did not receive prednisolone ($P=0.002$). **CONCLUSIONS:** Pentoxifylline did not improve survival in patients with alcoholic hepatitis. Prednisolone was associated with a reduction in 28-day mortality that did not reach significance and with no improvement in outcomes at 90 days or 1 year. (Funded by the National Institute for Health Research Health Technology Assessment program; STOPAH EudraCT number, 2009-013897-42, and Current Controlled Trials number, ISRCTN88782125.)

Question: Therapy

Study Type: RCT

Learner Level: Intermediate

Notes: Good article to discuss and explain odds ratios as they use them in this RCT (vs risk ratio).

Treanor, J. J., et al. (2000). "Efficacy and safety of the oral neuraminidase inhibitor oseltamivir in treating acute influenza: a randomized controlled trial. US Oral Neuraminidase Study Group." *JAMA* 283(8): 1016-1024.

CONTEXT: Previous studies have shown oseltamivir, a neuraminidase inhibitor, to be effective in preventing influenza and treating experimental influenza. OBJECTIVE: To evaluate the efficacy and safety of oseltamivir in the treatment of naturally acquired influenza infection. DESIGN: Randomized, placebo-controlled, double-blind study conducted January through March 1998. SETTING: Sixty primary care and university health centers throughout the United States. PARTICIPANTS: A total of 629 healthy nonimmunized adults aged 18 to 65 years with febrile respiratory illness of no more than 36 hours' duration with temperature of 38 degrees C or more plus at least 1 respiratory symptom and 1 constitutional symptom. INTERVENTIONS: Individuals were randomized to 1 of 3 treatment groups with identical appearing pills: oral oseltamivir phosphate, 75 mg twice daily (n = 211) or 150 mg (n = 209) twice daily, or placebo (n = 209). MAIN OUTCOME MEASURES: Duration and severity of illness in individuals infected with influenza. RESULTS: Two individuals withdrew before receiving medication and were excluded from further analyses. A total of 374 individuals (59.6%) were infected with influenza. Their duration of illness was reduced by more than 30% with both oseltamivir, 75 mg twice daily (median, 71.5 hours; $P < .001$), and oseltamivir, 150 mg twice daily (median, 69.9 hours; $P = .006$), compared with placebo (median, 103.3 hours). Severity of illness was reduced by 38% (median score, 597 score-hours; $P < .001$) with oseltamivir, 75 mg twice daily, and by 35% (median score, 626 score-hours; $P < .001$) with oseltamivir, 150 mg twice daily, vs placebo (median score, 963 score-hours). Oseltamivir treatment reduced the duration of fever and oseltamivir recipients returned to usual activities 2 to 3 days earlier than placebo recipients ($P < \text{or} = .05$). Secondary complications such as bronchitis and sinusitis occurred in 15% of placebo recipients compared with 7% of combined oseltamivir recipients ($P = .03$). Among all 629 subjects, oseltamivir reduced illness duration (76.3 hours and 74.3 hours for 75 mg and 150 mg, respectively, vs 97.0 hours for placebo; $P = .004$ for both comparisons) and illness severity (686 score-hours and 629 score-hours for 75 mg and 150 mg, respectively, vs 887 score-hours for placebo; $P < .001$ for both comparisons). Nausea and vomiting occurred more frequently in both oseltamivir groups (combined, 18.0% and 14.1%, respectively; $P = .002$) than in the placebo group (7.4% and 3.4%; $P < .001$). CONCLUSIONS: Our data suggest that oral oseltamivir treatment reduces the duration and severity of acute influenza in healthy adults and may decrease the incidence of secondary complications.

Question: Therapy

Study Type: RCT

Learner Level: Beginner

Notes: Fairly straightforward RCT; good for teaching basic principles of validity criteria. Topic good for discussion about whether a 2000 year old article still has relevance, given the world-wide epidemic of H1N1 influenza.

Turpie, A. G., et al. (2002). "Fondaparinux vs enoxaparin for the prevention of venous thromboembolism in major orthopedic surgery: a meta-analysis of 4 randomized double-blind studies." *Arch Intern Med* 162(16): 1833-1840.

BACKGROUND: Orthopedic surgery remains a condition at high risk of venous thromboembolism (VTE). Fondaparinux, the first of a new class of synthetic selective factor Xa inhibitors, may further reduce this risk compared with currently available thromboprophylactic treatments. METHODS: A meta-analysis of 4 multicenter, randomized, double-blind trials in patients undergoing elective hip replacement, elective major knee surgery, and surgery for hip fracture (N = 7344) was performed to determine whether a subcutaneous 2.5-mg, once-daily regimen of fondaparinux sodium starting 6 hours after surgery was more effective and as safe as approved enoxaparin regimens in preventing VTE. The primary efficacy outcome was VTE up to day 11, defined as deep vein thrombosis detected by mandatory bilateral venography or documented symptomatic deep vein thrombosis or pulmonary embolism. The primary safety outcome was major bleeding. RESULTS: Fondaparinux significantly reduced the incidence of VTE by day 11 (182 [6.8%] of 2682 patients) compared with enoxaparin (371 [13.7%] of 2703 patients), with a common odds reduction of 55.2% (95% confidence interval, 45.8% to 63.1%; $P < .001$); this beneficial effect was consistent across all types of surgery and all subgroups. Although major bleeding occurred more frequently in the fondaparinux-treated group ($P = .008$), the incidence of clinically relevant bleeding (leading to death or reoperation or occurring in a critical organ) did not differ between groups. CONCLUSION: In patients undergoing orthopedic surgery, 2.5 mg of fondaparinux sodium once daily, starting 6 hours postoperatively, showed a major benefit over enoxaparin, achieving an overall risk reduction of VTE greater than 50% without increasing the risk of clinically relevant bleeding.

Question: Therapy

Study Type: Meta-analysis

Learner Level: Beginner

Notes: Good to teach difference between a systematic review and a meta-analysis

Tyring, S., et al. (1995). "Famciclovir for the treatment of acute herpes zoster: effects on acute disease and postherpetic neuralgia. A randomized, double-blind, placebo-controlled trial. Collaborative Famciclovir Herpes Zoster Study Group." *Ann Intern Med* 123(2): 89-96.

OBJECTIVE: To document the effects of treatment with famciclovir on the acute signs and symptoms of herpes zoster and postherpetic neuralgia. DESIGN: A randomized, double-blind, placebo-controlled, multicenter trial. SETTING: 36 centers in the United States, Canada, and Australia. PATIENTS: 419 immunocompetent adults with uncomplicated herpes zoster. INTERVENTION: Patients were assigned within 72 hours of rash onset to famciclovir, 500 mg; famciclovir, 750 mg; or placebo, three times daily for 7 days. MEASUREMENTS: Lesions were assessed daily for as long as 14 days until full crusting occurred and then weekly until the lesions healed. Viral cultures were obtained daily while vesicles were present. Pain was assessed at each of the visits at which lesions were examined and then monthly for 5 months after the lesions healed. Safety was assessed throughout the study. RESULTS: Famciclovir was well tolerated, with a safety profile similar to that of placebo. Famciclovir accelerated lesion healing and reduced the duration of

viral shedding. Most importantly, famciclovir recipients had faster resolution of postherpetic neuralgia (approximately twofold faster) than placebo recipients; differences between the placebo group and both the 500-mg famciclovir group (hazard ratio, 1.7 [95% CI, 1.1 to 2.7]) and the 750-mg famciclovir group (hazard ratio, 1.9 [CI, 1.2 to 2.9]) were statistically significant ($P = 0.02$ and 0.01 , respectively). The median duration of postherpetic neuralgia was reduced by approximately 2 months. **CONCLUSIONS:** Oral famciclovir, 500 mg or 750 mg three times daily for 7 days, is an effective and well-tolerated therapy for herpes zoster that decreases the duration of the disease's most debilitating complication, postherpetic neuralgia.

Question: Therapy

Study Type: RCT

Learner Level: Beginner / Intermediate

Notes: Solid Methodology; Difficult to calculate NNT (number needed to treat)—but it can be done....Good discussion points: End points that are 'time' events; "intention to treat" vs. "efficacy-evaluable" analysis; Does it matter who paid for the trial? (See grant support page 95)

Umpierrez, G. E., et al. (2004). "Efficacy of subcutaneous insulin lispro versus continuous intravenous regular insulin for the treatment of patients with diabetic ketoacidosis." *Am J Med* 117(5): 291-296.

PURPOSE: To compare the efficacy and safety of subcutaneous insulin lispro with that of a standard low-dose intravenous infusion protocol of regular insulin in patients with uncomplicated diabetic ketoacidosis. **METHODS:** In this prospective, randomized open trial, 20 patients treated with subcutaneous insulin lispro were managed in regular medicine wards ($n=10$) or an intermediate care unit ($n=10$), while 20 patients treated with the intravenous protocol were managed in the intensive care unit. Patients treated with subcutaneous lispro received an initial injection of 0.3 unit/kg followed by 0.1 unit/kg/h until correction of hyperglycemia (blood glucose levels <250 mg/dL), followed by 0.05 to 0.1 unit/kg/h until resolution of diabetic ketoacidosis ($\text{pH} > \text{or} = 7.3$, bicarbonate $> \text{or} = 18$ mEq/L). Patients treated with intravenous regular insulin received an initial bolus of 0.1 unit/kg, followed by an infusion of 0.1 unit/kg/h until correction of hyperglycemia, then 0.05 to 0.1 unit/kg/h until resolution of diabetic ketoacidosis. **RESULTS:** Mean (\pm SD) admission biochemical parameters in patients treated with subcutaneous lispro (glucose: 674 ± 154 mg/dL; bicarbonate: 9.2 ± 4 mEq/L; pH : 7.17 ± 0.10) were similar to values in patients treated with intravenous insulin (glucose: 611 ± 264 mg/dL; bicarbonate: 10.6 ± 4 mEq/L; pH : 7.19 ± 0.08). The duration of treatment until correction of hyperglycemia (7 ± 3 hours vs. 7 ± 2 hours) and resolution of ketoacidosis (10 ± 3 hours vs. 11 ± 4 hours) in patients treated with subcutaneous lispro was not different than in patients treated with intravenous regular insulin. There were no deaths in either group, and there were no differences in the length of hospital stay, amount of insulin until resolution of diabetic ketoacidosis, or in the rate of hypoglycemia between treatment groups. Treatment of diabetic ketoacidosis in the intensive care unit was associated with 39% higher hospitalization charges than was treatment with subcutaneous lispro in a non-intensive care setting ($\$14,429 \pm \5243 vs. $\$8801 \pm \5549 , $P < 0.01$). **CONCLUSION:** Treatment of adult patients who have uncomplicated diabetic ketoacidosis with subcutaneous lispro every hour in a non-intensive care setting may be safe and more cost-effective than treatment with intravenous regular insulin in the intensive care unit.

Question: Therapy

Study Type: RCT

Learner Level: Intermediate-Advanced

Notes: Mainly good to teach about spin and confounders—industry funded trial in which two groups differ in many ways in addition to the difference in main intervention; Conclusion made by authors is not clearly supported by the experimental design or the data; Allocation concealment is unclear;

This paper is really best for advanced learners who are eager to learn how to untangle issues that contribute to 'spin' in industry funded research

Van Gelder, I. C., et al. (2010). "Lenient versus strict rate control in patients with atrial fibrillation." *N Engl J Med* 362(15): 1363-1373.

BACKGROUND: Rate control is often the therapy of choice for atrial fibrillation. Guidelines recommend strict rate control, but this is not based on clinical evidence. We hypothesized that lenient rate control is not inferior to strict rate control for preventing cardiovascular morbidity and mortality in patients with permanent atrial fibrillation. **METHODS:** We randomly assigned 614 patients with permanent atrial fibrillation to undergo a lenient rate-control strategy (resting heart rate <110 beats per minute) or a strict rate-control strategy (resting heart rate <80 beats per minute and heart rate during moderate exercise <110 beats per minute). The primary outcome was a composite of death from cardiovascular causes, hospitalization for heart failure, and stroke, systemic embolism, bleeding, and life-threatening arrhythmic events. The duration of follow-up was at least 2 years, with a maximum of 3 years. **RESULTS:** The estimated cumulative incidence of the primary outcome at 3 years was 12.9% in the lenient-control group and 14.9% in the strict-control group, with an absolute difference with respect to the lenient-control group of -2.0 percentage points (90% confidence interval, -7.6 to 3.5; $P < 0.001$ for the prespecified noninferiority margin). The frequencies of the components of the primary outcome were similar in the two groups. More patients in the lenient-control group met the heart-rate target or targets (304 [97.7%], vs. 203 [67.0%] in the strict-control group; $P < 0.001$) with fewer total visits (75 [median, 0], vs. 684 [median, 2]; $P < 0.001$). The frequencies of symptoms and adverse events were similar in the two groups. **CONCLUSIONS:** In patients with permanent atrial fibrillation, lenient rate control is as effective as strict rate control and is easier to achieve. (ClinicalTrials.gov number, NCT00392613.)

Question: Therapy

Study Type: RCT

Learner Level: Intermediate

Notes: Intermediate RCT studying treatment paradigm of strict versus lenient control for atrial fibrillation; Well designed study but with several topics worthy of discussion including composite outcomes. Very good for discussion of application of evidence – strict control was not reached in a significant portion of the group randomized to that arm. How does that effect interpretation? Article has an ACP-JC summary which can help teachers prepare to teach the paper and also illustrate how quick and useful such summaries can be.

Van Gelder, I. C., et al. (2002). "A comparison of rate control and rhythm control in patients with recurrent persistent atrial fibrillation." *N Engl J Med* 347(23): 1834-1840.

BACKGROUND: Maintenance of sinus rhythm is the main therapeutic goal in patients with atrial fibrillation. However, recurrences of atrial fibrillation and side effects of antiarrhythmic drugs offset the benefits of sinus rhythm. We hypothesized that ventricular rate control is not inferior to the maintenance of sinus rhythm for the treatment of atrial fibrillation. **METHODS:** We randomly assigned 522 patients who had persistent atrial fibrillation after a previous electrical cardioversion to receive treatment aimed at rate control or rhythm control. Patients in the rate-control group received oral anticoagulant drugs and rate-slowing medication. Patients in the rhythm-control group underwent serial cardioversions and received antiarrhythmic drugs and oral anticoagulant drugs. The end point was a composite of death from cardiovascular causes, heart failure, thromboembolic complications, bleeding, implantation of a pacemaker, and severe adverse effects of drugs. **RESULTS:** After a mean (+/-SD) of 2.3+/-0.6 years, 39 percent of the 266 patients in the rhythm-control group had sinus rhythm, as compared with 10 percent of the 256 patients in the rate-control group. The primary end point occurred in 44 patients (17.2 percent) in the rate-control group and in 60 (22.6 percent) in the rhythm-control group. The 90 percent (two-sided) upper boundary of the absolute difference in the primary end point was 0.4 percent (the prespecified criterion for noninferiority was 10 percent or less). The distribution of the various components of the primary end point was similar in the rate-control and rhythm-control groups. **CONCLUSIONS:** Rate control is not inferior to rhythm control for the prevention of death and morbidity from cardiovascular causes and may be appropriate therapy in patients with a recurrence of persistent atrial fibrillation after electrical cardioversion.

Question: Therapy

Study Type: RCT

Learner Level: Beginner

Notes: Straightforward RCT with clear methods and clear results reporting. Good Kaplan-Meier curves. Good easy numbers for calculation of number needed to treat (see table 2 which reports absolute risk differences). Also can use same table for discussion of confidence intervals and statistical significance. Good discussion points: Great data for discussion of primary vs. secondary end points and composite end points. Can be used to discuss the differences between non-inferiority / equivalence / superiority and considerations of sample size.

Velazquez, E. J., et al. (2019). "Angiotensin-Nepriylsin Inhibition in Acute Decompensated Heart Failure." *N Engl J Med* 380(6): 539-548.

BACKGROUND: Acute decompensated heart failure accounts for more than 1 million hospitalizations in the United States annually. Whether the initiation of sacubitril-valsartan therapy is safe and effective among patients who are hospitalized for acute decompensated heart failure is unknown. **METHODS:** We enrolled patients with heart failure with reduced ejection fraction who were hospitalized for acute decompensated heart failure at 129 sites in the United States. After hemodynamic stabilization, patients were randomly assigned to receive sacubitril-valsartan (target dose, 97 mg of sacubitril with 103 mg of valsartan twice daily) or enalapril (target dose, 10 mg twice daily). The primary efficacy outcome was the time-averaged proportional change in the N-terminal pro-B-type natriuretic peptide (NT-proBNP) concentration from baseline through weeks 4 and 8. Key safety outcomes were the rates of worsening renal function, hyperkalemia, symptomatic hypotension, and angioedema. **RESULTS:** Of the 881 patients who underwent randomization, 440 were assigned to receive sacubitril-valsartan and 441 to receive enalapril. The time-averaged reduction in the NT-proBNP concentration was significantly greater in the sacubitril-valsartan group than in the enalapril group; the ratio of the geometric mean of values obtained at weeks 4 and 8 to the baseline value was 0.53 in the sacubitril-valsartan group as compared with 0.75 in the enalapril group (percent change, -46.7% vs. -25.3%; ratio of change with sacubitril-valsartan vs. enalapril, 0.71; 95% confidence interval [CI], 0.63 to 0.81; P<0.001). The greater reduction in the NT-proBNP concentration with sacubitril-valsartan than with enalapril was evident as early as week 1 (ratio of change, 0.76; 95% CI, 0.69 to 0.85). The rates of worsening renal function, hyperkalemia, symptomatic hypotension, and angioedema did not differ significantly between the two groups. **CONCLUSIONS:** Among patients with heart failure with reduced ejection fraction who were hospitalized for acute decompensated heart failure, the initiation of sacubitril-valsartan therapy led to a greater reduction in the NT-proBNP concentration than enalapril therapy. Rates of worsening renal function, hyperkalemia, symptomatic hypotension, and angioedema did not differ significantly between the two groups. (Funded by Novartis; PIONEER-HF ClinicalTrials.gov number, NCT0254890 .).

Question: Therapy

Study Type: RCT

Learner Level: Beginner/Intermediate

Notes: This article is good for discussing whether the outcome measured is the one we care about and how much we can apply it. Also good for talking about bias as it promotes a drug and the study was funded by the pharmaceutical company that sells it.

Venkatesh, B., et al. (2018). "Adjunctive Glucocorticoid Therapy in Patients with Septic Shock." N Engl J Med 378(9): 797-808.

BACKGROUND: Whether hydrocortisone reduces mortality among patients with septic shock is unclear. **METHODS:** We randomly assigned patients with septic shock who were undergoing mechanical ventilation to receive hydrocortisone (at a dose of 200 mg per day) or placebo for 7 days or until death or discharge from the intensive care unit (ICU), whichever came first. The primary outcome was death from any cause at 90 days. **RESULTS:** From March 2013 through April 2017, a total of 3800 patients underwent randomization. Status with respect to the primary outcome was ascertained in 3658 patients (1832 of whom had been assigned to the hydrocortisone group and 1826 to the placebo group). At 90 days, 511 patients (27.9%) in the hydrocortisone group and 526 (28.8%) in the placebo group had died (odds ratio, 0.95; 95% confidence interval [CI], 0.82 to 1.10; $P=0.50$). The effect of the trial regimen was similar in six prespecified subgroups. Patients who had been assigned to receive hydrocortisone had faster resolution of shock than those assigned to the placebo group (median duration, 3 days [interquartile range, 2 to 5] vs. 4 days [interquartile range, 2 to 9]; hazard ratio, 1.32; 95% CI, 1.23 to 1.41; $P<0.001$). Patients in the hydrocortisone group had a shorter duration of the initial episode of mechanical ventilation than those in the placebo group (median, 6 days [interquartile range, 3 to 18] vs. 7 days [interquartile range, 3 to 24]; hazard ratio, 1.13; 95% CI, 1.05 to 1.22; $P<0.001$), but taking into account episodes of recurrence of ventilation, there were no significant differences in the number of days alive and free from mechanical ventilation. Fewer patients in the hydrocortisone group than in the placebo group received a blood transfusion (37.0% vs. 41.7%; odds ratio, 0.82; 95% CI, 0.72 to 0.94; $P=0.004$). There were no significant between-group differences with respect to mortality at 28 days, the rate of recurrence of shock, the number of days alive and out of the ICU, the number of days alive and out of the hospital, the recurrence of mechanical ventilation, the rate of renal-replacement therapy, and the incidence of new-onset bacteremia or fungemia. **CONCLUSIONS:** Among patients with septic shock undergoing mechanical ventilation, a continuous infusion of hydrocortisone did not result in lower 90-day mortality than placebo. (Funded by the National Health and Medical Research Council of Australia and others; ADRENAL ClinicalTrials.gov number, NCT01448109 .).

Question: Therapy

Study Type: RCT

Learner Level: Beginner

Notes: Well done and clearly reported RCT on a controversial topic (steroids in septic shock). Main outcome is no difference (90-day mortality), but other outcomes that are also important do show a difference (can be used to do calculations including NNT)

Walters, J. A., et al. (2009). "Systemic corticosteroids for acute exacerbations of chronic obstructive pulmonary disease." Cochrane Database Syst Rev(1): CD001288.

BACKGROUND: COPD is a common condition, mainly related to smoking. Acute exacerbations of COPD, usually related to superimposed infection, occur commonly and systemic corticosteroids are widely used in their management in combination with other treatments including antibiotics, oxygen supplementation and bronchodilators. **OBJECTIVES:** To determine the efficacy of corticosteroids, administered either parenterally or orally, on the outcomes of acute exacerbations of COPD. **SEARCH STRATEGY:** Searches were carried out using the Cochrane Airways Group COPD RCT register with additional studies sought in the bibliographies of randomised controlled trials and review articles. Authors of identified randomised controlled trials were contacted for other published and unpublished studies. The last search was carried out in August 2008. **SELECTION CRITERIA:** Randomised controlled trials comparing corticosteroids, administered either parenterally or orally, with appropriate placebo control. Other interventions e.g. bronchodilators and antibiotics were standardised. Clinical studies of acute asthma were excluded. **DATA COLLECTION AND ANALYSIS:** Data were extracted independently by two reviewers. Data measured but not reported were sought from authors of included studies. Trials were combined using Review Manager for analyses. **MAIN RESULTS:** Eleven studies ($n=1081$) fulfilled the inclusion criteria and 10 studies contributed data for analyses ($n=1051$). There were significantly fewer treatment failures within thirty days in patients given corticosteroid treatment, Odds Ratio (OR) 0.50; 95% confidence interval (CI) 0.36 to 0.69 and Hazard Ratio 0.78; 95% CI 0.63 to 0.97. It would have been necessary to treat 10 patients (95%CI 7 to 16) with corticosteroids to avoid one treatment failure in this time period. Duration of hospitalisation was significantly shorter with corticosteroid treatment, mean difference -1.22 days; 95% CI -2.26 to -0.18. For FEV1 there were significant treatment benefits with mean differences at the early time point (to 72 hours), 140 ml; 95% CI 90 to 190 ml and at end of treatment (up to 15 days) 80 ml; 95% confidence interval 10 to 160. There was a significant improvement in breathlessness and blood gases at both time points. There was no significant effect on mortality but an increased likelihood of an adverse event associated with corticosteroid treatment, OR 2.33; 95% CI 1.60 to 3.40. Overall one extra adverse effect occurred for every 5 people treated (95% CI 4 to 9). The risk of hyperglycaemia was significantly increased, OR 4.95; 95% CI 2.47 to 9.91. **AUTHORS' CONCLUSIONS:** Treatment of an exacerbation of COPD with oral or parenteral corticosteroids significantly reduces treatment failure and the need for additional medical treatment and shortens hospital stay. It increases the rate of improvement in lung function and dyspnoea and the improvement continues during treatment, but there is a significantly increased risk of an adverse drug event occurring. The optimal dose and length of treatment regime needs to be better defined.

Question: Therapy

Study Type: Meta-analysis

Learner Level: Intermediate

Notes: Good article for review of the critical appraisal of a systematic review as methods are clearly reported. Teaching point can be made when discussing the outcome of 'treatment failure'. Page 8 describes the different meanings for this composite outcome in the various studies. The question to focus on: are the various outcomes in the composite measure similar in importance to the patient? (all cause mortality, readmission, intensification of pharmacologic treatment). Forest plots are clear and do have plots with both significant heterogeneity (figure 2 I² = 43%) and others with I² = 0.

Winblad, B., et al. (2001). "A 1-year, randomized, placebo-controlled study of donepezil in patients with mild to moderate AD." *Neurology* 57(3): 489-495.

OBJECTIVE: To evaluate the long-term clinical efficacy and safety of donepezil versus placebo over 1 year in patients with mild to moderate AD. **METHODS:** Patients (n = 286; mean age, 72.5 years) with possible or probable AD from five Northern European countries were randomized to receive either donepezil (n = 142; 5 mg/day for 28 days, followed by 10 mg/day) or placebo (n = 144) for 1 year. **RESULTS:** The study was completed by 66.9% of the donepezil- and 67.4% of the placebo-treated patients. The benefit of donepezil over placebo was demonstrated by the Gottfries-Brane-Steen (a global assessment for rating dementia symptoms) total score at weeks 24, 36, and 52 (p < 0.05) and at the study end point (week 52, last observation carried forward; p = 0.054). Advantages of donepezil over placebo were also observed in cognition and activities of daily living (ADL) assessed by the Mini-Mental State Examination at weeks 24, 36, and 52, and the end point (p < 0.02) and by the Progressive Deterioration Scale at week 52 and the end point (p < 0.05). Adverse events (AE) were recorded for 81.7% of donepezil- and 75.7% of placebo-treated patients, with 7% of donepezil- and 6.3% of placebo-treated patients discontinuing because of AE. Treatment response to donepezil was not predicted by APOE genotype or sex in this population. **CONCLUSION:** As the first 1-year, multinational, double-blinded, placebo-controlled study of a cholinesterase inhibitor in AD, these data support donepezil as a well tolerated and effective long-term treatment for patients with AD, with benefits over placebo on global assessment, cognition, and ADL.

Question: Therapy

Study Type: RCT

Learner Level: Beginner / Intermediate

Notes: Strong RCT methodology. Good discussion points: Paper is good for calculations of risk ratio, absolute risk reduction and number needed to treat.

Wright, J. M. and V. M. Musini (2009). "First-line drugs for hypertension." *Cochrane Database Syst Rev*(3): CD001841.

BACKGROUND: Sustained elevated blood pressure, unresponsive to lifestyle measures, leads to a critically important clinical question: What class of drug to use first-line? This review answers that question. **OBJECTIVES:** Primary objective: To quantify the benefits and harms of the major first-line anti-hypertensive drug classes: thiazides, beta-blockers, calcium channel blockers, angiotensin converting enzyme (ACE) inhibitors, alpha-blockers, and angiotensin II receptor blockers (ARB). **SEARCH STRATEGY:** Electronic search of MEDLINE (Jan. 1966-June 2008), EMBASE, CINAHL, the Cochrane clinical trial register, using standard search strategy of the hypertension review group with additional terms. **SELECTION CRITERIA:** Randomized trials of at least one year duration comparing one of 6 major drug classes with a placebo or no treatment. More than 70% of people must have BP >140/90 mmHg at baseline. **DATA COLLECTION AND ANALYSIS:** The outcomes assessed were mortality, stroke, coronary heart disease (CHD), cardiovascular events (CVS), decrease in systolic and diastolic blood pressure, and withdrawals due to adverse drug effects. Risk ratio (RR) and a fixed effects model were used to combine outcomes across trials. **MAIN RESULTS:** Of 57 trials identified, 24 trials with 28 arms, including 58,040 patients met the inclusion criteria. Thiazides (19 RCTs) reduced mortality (RR 0.89, 95% CI 0.83, 0.96), stroke (RR 0.63, 95% CI 0.57, 0.71), CHD (RR 0.84, 95% CI 0.75, 0.95) and CVS (RR 0.70, 95% CI 0.66, 0.76). Low-dose thiazides (8 RCTs) reduced CHD (RR 0.72, 95% CI 0.61, 0.84), but high-dose thiazides (11 RCTs) did not (RR 1.01, 95% CI 0.85, 1.20). Beta-blockers (5 RCTs) reduced stroke (RR 0.83, 95% CI 0.72, 0.97) and CVS (RR 0.89, 95% CI 0.81, 0.98) but not CHD (RR 0.90, 95% CI 0.78, 1.03) or mortality (RR 0.96, 95% CI 0.86, 1.07). ACE inhibitors (3 RCTs) reduced mortality (RR 0.83, 95% CI 0.72-0.95), stroke (RR 0.65, 95% CI 0.52-0.82), CHD (RR 0.81, 95% CI 0.70-0.94) and CVS (RR 0.76, 95% CI 0.67-0.85). Calcium-channel blocker (1 RCT) reduced stroke (RR 0.58, 95% CI 0.41, 0.84) and CVS (RR 0.71, 95% CI 0.57, 0.87) but not CHD (RR 0.77 95% CI 0.55, 1.09) or mortality (RR 0.86 95% CI 0.68, 1.09). No RCTs were found for ARBs or alpha-blockers. **AUTHORS' CONCLUSIONS:** First-line low-dose thiazides reduce all morbidity and mortality outcomes. First-line ACE inhibitors and calcium channel blockers may be similarly effective but the evidence is less robust. First-line high-dose thiazides and first-line beta-blockers are inferior to first-line low-dose thiazides.

Question: Therapy

Study Type: Meta-analysis

Learner Level: Intermediate to Advanced

Notes: This is an extensive review of first line therapy for hypertension including studies comparing major drug classes with placebo/ no treatment looking at outcomes of mortality, stroke and coronary artery disease. Although this meta-analysis is very long (60 pages), a teaching session could focus on particular parts, such as specific forest plots. There is also an ACP journal club summary that could serve as a central point of discussion. The key messages here are that there are the gaps between what we know (thiazides and ACE inhibitors decrease mortality) and what we do (a minority of patients with hypertension are on these drugs as first line therapies).

Wright, J. T., Jr., et al. (2015). "A Randomized Trial of Intensive versus Standard Blood-Pressure Control." *N Engl J Med* 373(22): 2103-2116.

BACKGROUND: The most appropriate targets for systolic blood pressure to reduce cardiovascular morbidity and mortality among persons without diabetes remain uncertain. **METHODS:** We randomly assigned 9361 persons with a systolic blood pressure of 130 mm Hg or higher and an increased cardiovascular risk, but without diabetes, to a systolic blood-pressure target of less than 120 mm Hg (intensive treatment) or a target of less than 140 mm Hg (standard treatment). The primary composite outcome was myocardial infarction, other acute coronary syndromes, stroke, heart failure, or death from cardiovascular causes. **RESULTS:** At 1 year, the mean systolic blood pressure was 121.4 mm Hg in the intensive-treatment group and 136.2 mm Hg in the standard-treatment group. The intervention was stopped early after a median follow-up of 3.26 years owing to a significantly lower rate of the primary composite outcome in the intensive-treatment group than in the standard-treatment group (1.65% per year vs. 2.19% per year; hazard ratio with intensive treatment, 0.75; 95% confidence interval [CI], 0.64 to 0.89; $P < 0.001$). All-cause mortality was also significantly lower in the intensive-treatment group (hazard ratio, 0.73; 95% CI, 0.60 to 0.90; $P = 0.003$). Rates of serious adverse events of hypotension, syncope, electrolyte abnormalities, and acute kidney injury or failure, but not of injurious falls, were higher in the intensive-treatment group than in the standard-treatment group. **CONCLUSIONS:** Among patients at high risk for cardiovascular events but without diabetes, targeting a systolic blood pressure of less than 120 mm Hg, as compared with less than 140 mm Hg, resulted in lower rates of fatal and nonfatal major cardiovascular events and death from any cause, although significantly higher rates of some adverse events were observed in the intensive-treatment group. (Funded by the National Institutes of Health; ClinicalTrials.gov number, NCT01206062.)

Question: Therapy

Study Type: RCT

Learner Level: Beginner

Notes: Teaching points: high profile study (reported in NY Times before the study was analyzed) about common problem (Hypertension). Results are at odds with current CPG. Methodologic points: overall strong methods but with a few tricky points: trial was stopped early for benefit; good to talk about blinding (how do you handle in a strategy trial?), important applicability questions (to whom should this apply?)

Yeh, R. W., et al. (2018). "Parachute use to prevent death and major trauma when jumping from aircraft: randomized controlled trial." BMJ 363: k5094.

OBJECTIVE: To determine if using a parachute prevents death or major traumatic injury when jumping from an aircraft. **DESIGN:** Randomized controlled trial. **SETTING:** Private or commercial aircraft between September 2017 and August 2018. **PARTICIPANTS:** 92 aircraft passengers aged 18 and over were screened for participation. 23 agreed to be enrolled and were randomized. **INTERVENTION:** Jumping from an aircraft (airplane or helicopter) with a parachute versus an empty backpack (unblinded). **MAIN OUTCOME MEASURES:** Composite of death or major traumatic injury (defined by an Injury Severity Score over 15) upon impact with the ground measured immediately after landing. **RESULTS:** Parachute use did not significantly reduce death or major injury (0% for parachute v 0% for control; $P > 0.9$). This finding was consistent across multiple subgroups. Compared with individuals screened but not enrolled, participants included in the study were on aircraft at significantly lower altitude (mean of 0.6 m for participants v mean of 9146 m for non-participants; $P < 0.001$) and lower velocity (mean of 0 km/h v mean of 800 km/h; $P < 0.001$). **CONCLUSIONS:** Parachute use did not reduce death or major traumatic injury when jumping from aircraft in the first randomized evaluation of this intervention. However, the trial was only able to enroll participants on small stationary aircraft on the ground, suggesting cautious extrapolation to high altitude jumps. When beliefs regarding the effectiveness of an intervention exist in the community, randomized trials might selectively enroll individuals with a lower perceived likelihood of benefit, thus diminishing the applicability of the results to clinical practice.

Question: Therapy

Study Type: RCT

Learner Level: Beginner

Notes: Fun topic. Use it to introduce alidity criteria (assessment of risk of bias).

Zanasi, A., et al. (2015). "Does additional antimicrobial treatment have a better effect on URTI cough resolution than homeopathic symptomatic therapy alone? A real-life preliminary observational study in a pediatric population." Multidiscip Respir Med 10(1): 25.

BACKGROUND: The effectiveness of a homeopathic syrup on cough has been demonstrated in an adult population in a previous double-blind randomized study. The present prospective observational study investigated children affected by wet acute cough caused by non-complicated URTIs, comparing those who received the homeopathic syrup versus those treated with the homeopathic syrup plus antibiotic. **OBJECTIVES:** The aims were: 1) to assess whether the addition of antibiotics to a symptomatic treatment had a role in reducing the severity and duration of acute cough in a pediatric population, as well as in improving cough resolution; 2) to verify the safety of the two treatments. **METHODS:** Eighty-five children were enrolled in an open study: 46 children received homeopathic syrup alone for 10 days and 39 children received homeopathic syrup for 10 days plus oral antibiotic treatment (amoxicillin/clavulanate, clarithromycin, and erythromycin) for 7 days. To assess cough severity we used a subjective verbal category-descriptive (VCD) scale. **RESULTS:** Cough VCD score was significantly ($P < 0.001$) reduced in both groups starting from the second day of treatment (-0.52 +/- 0.66 in the homeopathic syrup group and -0.56 +/- 0.55 in children receiving homeopathic syrup plus oral antibiotic treatment). No significant differences in cough severity or resolution were found between the two groups of children in any of the 28 days of the study. After the first week (day 8) cough was completely resolved in more than one-half of

patients in both groups. Two children (4.3 %) reported adverse effects in the group treated with the homeopathic syrup alone, versus 9 children (23.1 %) in the group treated with the homeopathic syrup plus antibiotics (P = 0.020). CONCLUSIONS: Our data confirm that the homeopathic treatment in question has potential benefits for cough in children as well, and highlight the strong safety profile of this treatment. Additional antibiotic prescription was not associated with a greater cough reduction, and presented more adverse events than the homeopathic syrup alone.

Question: Therapy

Study Type: Prospective cohort study

Learner Level: Beginner

Notes: Use this article to teach bias and general critical appraisal, as the whole article is pretty bad – comparing kids on a homeopathic syrup to a group of kids on the syrup plus antibiotic for the common cold, and looking at which group suffered more side effects. The lead author can be seen on the syrup's website giving an attestation of how well it works. Conflict of interest!

Zinman, B., et al. (2015). "Empagliflozin, Cardiovascular Outcomes, and Mortality in Type 2 Diabetes." *N Engl J Med* 373(22): 2117-2128.

BACKGROUND: The effects of empagliflozin, an inhibitor of sodium-glucose cotransporter 2, in addition to standard care, on cardiovascular morbidity and mortality in patients with type 2 diabetes at high cardiovascular risk are not known. **METHODS:** We randomly assigned patients to receive 10 mg or 25 mg of empagliflozin or placebo once daily. The primary composite outcome was death from cardiovascular causes, nonfatal myocardial infarction, or nonfatal stroke, as analyzed in the pooled empagliflozin group versus the placebo group. The key secondary composite outcome was the primary outcome plus hospitalization for unstable angina. **RESULTS:** A total of 7020 patients were treated (median observation time, 3.1 years). The primary outcome occurred in 490 of 4687 patients (10.5%) in the pooled empagliflozin group and in 282 of 2333 patients (12.1%) in the placebo group (hazard ratio in the empagliflozin group, 0.86; 95.02% confidence interval, 0.74 to 0.99; P=0.04 for superiority). There were no significant between-group differences in the rates of myocardial infarction or stroke, but in the empagliflozin group there were significantly lower rates of death from cardiovascular causes (3.7%, vs. 5.9% in the placebo group; 38% relative risk reduction), hospitalization for heart failure (2.7% and 4.1%, respectively; 35% relative risk reduction), and death from any cause (5.7% and 8.3%, respectively; 32% relative risk reduction). There was no significant between-group difference in the key secondary outcome (P=0.08 for superiority). Among patients receiving empagliflozin, there was an increased rate of genital infection but no increase in other adverse events. **CONCLUSIONS:** Patients with type 2 diabetes at high risk for cardiovascular events who received empagliflozin, as compared with placebo, had a lower rate of the primary composite cardiovascular outcome and of death from any cause when the study drug was added to standard care. (Funded by Boehringer Ingelheim and Eli Lilly; EMPA-REG OUTCOME ClinicalTrials.gov number, NCT01131676.).

Question: Therapy

Study Type: RCT

Learner Level: Beginner/Intermediate

Notes: Good study to address funding bias. Also, this study was conducted both as a non-inferiority trial and a more typical superiority RCT. Starting in 2008 the FDA required drug companies to prove cardiovascular non-inferiority for new diabetes medications. Of note, they performed a sensitivity analysis of death from any cause in which they assumed that all patients lost to follow up in the empagliflozin group died and all those lost in the placebo arm lived, and empagliflozin still showed a significant benefit. Note, however, that they did NOT perform (or at least report) this sensitivity analysis for the primary outcome, for which the study was actually powered and which had a confidence interval VERY close to being non-significant. Modified intention to treat...which is, of course, not intention to treat.