

Diagnosis – Teaching Papers

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.

Type of Question: Screening **Intermediate / Advanced Prospective cohort.**

Teaching 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....

Abstract: **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.

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

Type of Question: Screening **Intermediate Non-randomized, evaluator blinded, non-inferiority study.**

Teaching 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.

Abstract: **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.

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

Type of Question: Screening **Advanced RCT.**

Teaching 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

Abstract: 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.)

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

Type of Question: Screening **RCT.**

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

Abstract: 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.)

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

Type of Question: Screening **Cohort study.**

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

Abstract: **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.

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

Type of Question: Screening **Beginner**

Teaching 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.

Abstract: **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

Duke EBM Workshop – Teaching and Leading EBM March 2013

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.

Schneider, H. G., L. Lam, 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.

Type of Question: Diagnostic testing as intervention **Intermediate**

Teaching 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)

Abstract: 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.

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

Type of Question: Diagnosis **Intermediate / Advanced Meta analysis.**

Teaching 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.

Abstract: 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.

Collins, J. F., D. A. Lieberman, 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.

Type of Question: Diagnosis **Beginner / Intermediate Prospective cohort.**

Teaching 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?

Abstract: 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 Duke EBM Workshop – Teaching and Leading EBM

March 2013

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.

Mueller, C., A. Scholer, 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.

Type of Question: Diagnosis **Intermediate RCT of a testing strategy.**

Teaching 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?

Abstract: 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.

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

Type of Question: Diagnosis **Intermediate / Advanced Meta analysis.**

Teaching 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 Duke EBM Workshop – Teaching and Leading EBM

March 2013

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.

Abstract: 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.

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

Type of Question: Diagnosis **Beginner / Intermediate Prospective cohort.**

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

Abstract: 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.

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

Type of Question: Diagnosis **Intermediate / Advanced Prospective cohort.**

Teaching Notes: Okay... This is for those of you who object to PIOPED 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.

Abstract: 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.

Rathbun, S. W., G. E. Raskob, 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.

Type of Question: Diagnosis **Intermediate / Advanced Systematic review / Meta analysis.**

Teaching 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 Med* 2001;125(2):88-97 for example of an original study

Abstract: 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.

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

Type of Question: Diagnosis **Beginner / Intermediate Prospective cohort.**

Teaching Notes: Clear, careful methodology well described; Good discussion points; Reference Standards Inter-observer agreement (kappa); Straightforward data for calculation of Likelihood ratios

Abstract: 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.

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.

Type of Question: Diagnosis **Beginner / Intermediate Systematic review / Meta analysis.**

Teaching 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

Abstract: 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.

Grayson, M. L., G. W. Gibbons, et al. (1995). "**Probing to bone in infected pedal ulcers. A clinical sign of underlying osteomyelitis in diabetic patients.**" *Jama* 273(9): 721-723.

Type of Question: Diagnosis **Beginner / Intermediate Prospective cohort.**

Teaching 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)

Abstract: **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.

(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.

Type of Question: Diagnosis **Beginner / Intermediate; Classic paper Prospective cohort.**

Teaching 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

Abstract: 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.

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

Type of Question: Diagnosis **Beginner / Intermediate; Classic paper Prospective cohort.**

Teaching 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

Abstract: 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%.

Baarslag, H. J., E. J. van Beek, 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.

Type of Question: Diagnosis **Intermediate Prospective cohort.**

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

Abstract: 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.

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

Type of Question: Diagnosis **Intermediate Cohort.**

Teaching 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.

Abstract: 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.

Chung, T., A. Sindone, 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.

Type of Question: Diagnosis **Beginner**

Teaching Notes: Don't present likelihood ratios themselves but can get from data; Great for discussion of how LRs can differ when target negative population different

Abstract: 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.

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

Type of Question: Diagnosis **Beginner**

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

Abstract: 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.

Bush, K., D. R. Kivlahan, 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.

Type of Question: Diagnosis **Intermediate**

Teaching 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

Abstract: 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.

Kuppermann, N., J. F. Holmes, 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.

Type of Question: Diagnosis **Intermediate Clinical Prediction Rule.**

Teaching 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.

Abstract: 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 (ciTBI) 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 ciTBI (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%); ciTBIs 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 ciTBI 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 ciTBIs 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.

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

Type of Question: Diagnosis **Intermediate Clinical Prediction Rule.**

Teaching 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.

Abstract: 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 <or=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.

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

Type of Question: Diagnosis **Advanced Meta-analysis.**

Teaching Notes: Good to teach likelihood ratios

Abstract: 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.

Pouplard, C., P. Gueret, 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.

Type of Question: Diagnosis **Beginner Propsective cohort.**

Teaching Notes: Good to teach likelihood ratios and clinical prediction rules

Abstract: 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.

Colli, A., M. Fraquelli, 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.

Type of Question: Diagnosis **Advanced Systematic review.**

Teaching 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

Abstract: 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.

Chung, T., A. Sindone, 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.

Type of Question: Diagnosis

Teaching 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.

Abstract: 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.

Budoff, M. J., D. Dowe, 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.

Type of Question: Diagnosis **Intermediate Diagnostic Tests.**

Teaching 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

Abstract: **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).