Teaching and Leading Evidence-Based Practice

“Therapy”

Ken Goldberg, M.D.
Durham VA Healthcare System
Duke University Medical Center
April 28th, 2021
Goals and Disclaimers

• Review validity criteria for a therapy trial
• Review selected calculations and quantitative concepts related to therapy trials (i.e., “therapy math”)
• Apply the above to an actual article

*I have shamelessly stolen material from others – thanks in particular to Sheri Keitz, Daniella Zipkin, Denise Campbell-Sherer, Dr. Google, and EBM enthusiasts of all kinds

** Sadly, financial conflicts are few and far between in evidence-based medicine. I have none to disclose, but I am open to further discussion on this topic.

*** All jokes are the responsibility of the presenter in his capacity as a private citizen. Officially, the Department of Veterans Affairs has no sense of humor.
Why formalize this?
Categorization of reported outcomes

• **True** positive – there really is an effect (that I care about)

• **True** negative – there really is not an effect (or at least not one I care about)

• **False** Positive
  • Alpha ($\alpha$) error, or “false Alarm”

• **False** Negative
  • Beta ($\beta$) error, “missing the Boat”
What if we don’t?

Internet Searches for Unproven COVID-19 Therapies in the United States
Michael Liu, AB; Theodore L. Caputi, MPH; Mark Dredze, PhD; et al

$1 Million in Toxic Bleach Sold as ‘Miracle’ Cure, Officials Say
A Florida grand jury indicted a man and his three sons, accusing them of marketing an “unproven and potentially harmful” solution as a remedy for Covid-19, cancer, autism and more.

The Covid-19 Plasma Boom Is Over. What Did We Learn From It?
The U.S. government invested $800 million in plasma when the country was desperate for Covid-19 treatments. A year later, the program has fizzled.
By Katie Thomas and Noah Weiland, April 17, 2021
And now, a (mostly) true story
The Carolina Reaper is a cultivar of the Capsicum chinense plant. Developed by South Carolina breeder Ed Currie, the pepper is red and gnarled, with a bumpy texture and small pointed tail. In 2017, Guinness World Records declared it the hottest chili pepper in the world, surpassing the previous record holder, the Trinidad Scorpion "Butch T".


Lesson: “Do the research!”
Back to the story ...
What would you be worried about?

(Interactive – welcome to Sli.do!)

Join at slido.com: #EBMtherapy2021

https://app.sli.do/event/ib6ewdiw
Things I was worried about ...

• Lifetime number of sternotomies (zero preferred)
• Cumulative exposure to serious anticoagulation (none preferred)
  • Bleeding
  • Clotting
  • On the plus side, no more kale
• Valve “issues” (none preferred)
  • Mechanical failure
  • Infection
Kicking and screaming -

• Something is going to be “done” to me (i.e., a therapy)
• “Don’t get mad, get data” (S. Keitz)
• Which study design is best able to address my concerns (which is to say, reliably generalize to my situation / predict my future)?
  • Case Series
  • Case Control
  • Cohort
  • Randomized Control Trial
  • Meta-analysis
Anatomy of an RCT (1)
Validity Criteria for a therapy trial

• Did the intervention and control groups start with the same prognosis?
  • Were patients randomized?
  • Was randomization concealed?
  • Were patients in the study groups similar with respect to known prognostic factors?

• Was prognostic balance maintained as the study progressed?
  • To what extent was the study blinded?
  • Aside from the experimental intervention, were the groups treated equally?

• Were the groups prognostically balanced at the study' s completion?
  • Was follow-up complete?
  • Were patients analyzed in the groups to which they were randomized? (i.e., was it an intention-to-treat analysis?)
  • Was the trial stopped early?
Validity Criteria for a therapy trial, continued

• What are the results?
  • How precise was the estimate of the treatment effect?
  • How large was the treatment effect?

• How can I apply the results to patient care?
  • Were the study patients similar to my patient?
  • Were all patient-important outcomes considered?
  • Are the likely treatment benefits worth the potential harms and costs?
Anatomy of an RCT (2)
Prognosis at the beginning

• Randomization

• Allocation concealment

• Similar at baseline
Randomization

• Why randomize?
  • Spread confounding variables evenly across the groups
  • Increase likelihood that the intervention is the only difference between groups
Allocation Concealment

• Investigator cannot influence the allocation at the time of study entry

• RCTs lacking a statement about allocation concealment are associated with larger effect-size bias (33% if unclear, 41% if not done)

• Not typically done (55% of RCTs in “best” journals, 7% of RCTs in “poorer” journals)

Similar at Baseline

- Known prognostic factors should be balanced between groups.
  - Typically, “Table 1”

- Discussion point: Are $p$-values meaningful in Table 1?
Prognosis in the middle

• Blinding

• Equal treatment
Blinding

• Decreases the risk of biased assessments
• Whom shall we blind?
  • Subjects
  • Clinicians
  • Data collectors
  • Outcome adjudicators
  • Data analysts
Blinding

• For questions of stroke or bleeding, where might bias come into play if adjudicators are not blinded to treatment assignment?
Equal Treatment

• The experimental intervention should be the only thing that differs between groups.

• Any other factor which differs systematically between groups is called a “co-intervention” and may obscure true results.
Prognosis at the end

• Follow up complete

• Intention-to-treat analysis

• Trials stopped early
Follow-up Complete

• Loss of subjects creates missing data, which threatens the balance of randomization

• Those lost may have different prognosis than those who stayed

• Methods for managing missing data vary in strength
Intention to Treat (ITT)

• Analyzed in the groups to which they were randomized (even if they didn’t get the intervention)

• Cross-over introduces bias
  • A bias in which direction?

• ITT = “Effectiveness”

• Euphemisms for breaking ITT: “per protocol analysis”, “as treated analysis”, “efficacy analysis”
Not Intention to Treat
Trials Stopped Early

• Fewer observed outcomes
• Greater chance of random error

• Truncating RCTs accounts for differences in effect size, in a systematic review
• Magnitude is greatest with fewer than 500 outcome events

Bassler et al. JAMA 2010;303(12):1180-1187
Once you determine the paper is worth reading ...

Therapy math

Don’t worry – we have cookies!
Break (5 minutes)!

• Mmm... cookies ...
Once you determine the paper is worth reading ...

Therapy math
How are results presented?

- **Risk** = events or outcomes
- **Absolute Risk** = proportion of group with an outcome = event rate
- For (a purely fictitious) example – from a population of mechanical aortic valve replacements patients:
  - 100 patients were randomized to drug regimen A. At 1 year, 15 patients threw a clot.
  - 100 patients were randomized to drug regimen B. At 1 year, 10 patients threw a clot.
Determination of “Comparative Risk”

- **Drug A**: event rate is 15% (15 out of 100)

- **Drug B**: event rate is 10% (10 out of 100)

- There are only two things we can (or should) do to these numbers:
  - Subtract them
  - Divide them
Subtract

• 15% **MINUS** 10% = 5%

15\% - 10\% = 5\%

• What is this number (5%)?

• Absolute Risk Reduction (ARR)
Divide

- 10% DIVIDED BY 15% = 0.67%

- What do we call this number (0.67%)?
  - Relative Risk (RR)
  - Risk Ratio
Relative Risk Reduction (1)

- **RRR = 1 − RR**

And...
- **RRR = ARR/baseline risk**

- **RRR = 1 − 0.67 = 0.33 = 33%**
Relative Risk Reduction (2)

• $\text{RRR} = \frac{\text{ARR}}{\text{baseline risk}}$

• “What proportion of our baseline risk have we reduced?”

• $\text{RRR} (100\% - 10\%/15\% = 5\%/15\%) = 33\%$

• The $\text{ARR} (15\%-10\%)$ was 5%

• Which number is subjectively more impressive (and thus more likely to be reported, especially if you use the adjective “better”)?
Absolute Risk Reduction (ARR) vs. Relative Risk Reduction (RRR)

ARR: 25%

RRR: 50%

50%
Number Needed to Treat (NNT)

NNT = 1/ARR

If ARR = 5%...

Then treating 100 people changes the outcome in 5

How many people must I treat to change the outcome in 1?

100  5
20   1
And the valve shall be On-X...

Anticoagulation and Antiplatelet Strategies After On-X Mechanical Aortic Valve Replacement

John D. Puskas, MD, MSc, a Marc Gerdisch, MD, b Dennis Nichols, MD, c Lilibeth Fermin, MD, d Birger Rhenman, MD, d Divya Kapoor, MD, d Jack Copeland, MD, e Reed Quinn, MD, f G. Chad Hughes, MD, g Hormoz Azar, MD, h Michael McGrath, MD, h Michael Wait, MD, i Bobby Kong, MD, i Tomas Martin, MD, i E. Charles Douville, MD, i Steven Meyer, MD, PhD, m Jian Ye, MD MSc, n W.R. Eric Jamieson, MD, o Lance Landvater, MD, p Robert Hagberg, MD, q Timothy Trotter, MD, r John Armitage, MD, s Jeffrey Askew, MD, t Kevin Accola, MD, j Paul Levy, MD, u David Duncan, MD, v Bobby Yanagawa, MD, PhD, w John Ely, MS, x Allen Graeve, MD, c for the PROACT Investigators*
Let’s apply the validity criteria

• How were they randomized? (A: Groups 1-3)
• Same at baseline? (B: Groups 4-6)
• How was blinding handled during the trial? (C: Groups 7-9)
• Was follow-up complete? (D: Groups 10-12)
• Intention-to-treat analysis? (E: Groups 13-16)
• Was the trial stopped early? (Also E: Groups 13-16)

• Bonus question: How applicable are results to me (it’s all about me)? (all – Group X)
Things to ignore

• The word “non-inferiority”
• This is actually more than one trial
  • One recruitment pool, but independent interventions
  • We will focus on only one intervention
• Survival analysis (including Kaplan-Meier references)
Break into groups, address assigned criteria

Go!
Randomized?

• (A: Groups 1-3)

**RANDOMIZATION AND MASKING.** Patients were randomly assigned (1:1) to intervention or a standard therapy control group via a secure Web-based central randomization system.
Same at baseline?

- (B: Groups 4-6)
Blinded?

- (C: Groups 7-9)

- Any information on the following?
  - Subjects
  - Clinicians
  - Data collectors
  - Outcome adjudicators
  - Data analysts

The Clinical Events Committee was masked to group assignment while adjudicating events.
Was follow-up complete?

- (D: Groups 10-12)
Intention to treat (ITT)?

• (E: Groups 13-16)

All analyses were conducted according to the intention-to-treat principle.
Was the trial stopped early?

- (F: Groups 13-16)

Enrollment was closed in January 2014 when the Data and Safety Monitoring Board recommended to the Steering Committee and to the study sponsor that enrollment in the low-risk arm of the PROACT trial be terminated because of increased cerebral TE events in the treatment group.
Results

• I am intentionally asking for a different way of looking at the results, based around things I care about (remember, this is hyper-personal)
  • I only care about major bleeding or stroke / stroke-adjacent events
  • Don’t worry about p-values

• We will focus event rate measured in events / enrollee, not events per patient-year as reported
Results

• Dual anti-platelet group (Groups 1-8)
  • How many people were in this group?
  • How many major bleeding and neurological events did they experience?
  • What is the “event rate” for them?

• How many were in warfarin group? (Groups 9-16)
  • How many people were in this group?
  • How many major bleeding neurological events did they experience?
  • What is the event rate for them?
Groups A and B
## Results

<table>
<thead>
<tr>
<th></th>
<th>Standard Warfarin (INR 2.0-3.0) (343.5 pt-ys)</th>
<th>DAPT (288.1 pt-ys)</th>
<th>Rate Ratio (DAPT/Standard-Dose Warfarin)</th>
<th>95% CI</th>
<th>p Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Primary endpoint</td>
<td>13</td>
<td>29</td>
<td>2.66</td>
<td>1.38-5.12</td>
<td>0.003</td>
</tr>
<tr>
<td>Components of co-primary endpoint</td>
<td>9 (Major bleeding)</td>
<td>2 (Major bleeding)</td>
<td>0.79</td>
<td>0.28-2.23</td>
<td>0.70</td>
</tr>
<tr>
<td>Major bleeding</td>
<td>0.26</td>
<td>0.00</td>
<td>0</td>
<td>0.47-8.32</td>
<td>0.30</td>
</tr>
<tr>
<td>Cerebral bleeding</td>
<td>0.87</td>
<td>5.74</td>
<td>1.99</td>
<td>0.48-2.48</td>
<td>0.80</td>
</tr>
<tr>
<td>Minor bleeding</td>
<td>12</td>
<td>11</td>
<td>1.09</td>
<td>0.22-26.3</td>
<td>0.50</td>
</tr>
<tr>
<td>Total bleeding</td>
<td>3.49</td>
<td>3.82</td>
<td>2.38</td>
<td>1.36-84.7</td>
<td>0.02</td>
</tr>
<tr>
<td>Stroke</td>
<td>0.00</td>
<td>7.24</td>
<td>16.69</td>
<td>2.20-127</td>
<td>0.007</td>
</tr>
<tr>
<td>TIA</td>
<td>0.29</td>
<td>0.00</td>
<td>0</td>
<td>1.37-5.98</td>
<td>0.005</td>
</tr>
<tr>
<td>Any neurological event</td>
<td>1.29</td>
<td>9.12</td>
<td>10.73</td>
<td>0.07-19.06</td>
<td>0.90</td>
</tr>
<tr>
<td>Peripheral TE event</td>
<td>0.00</td>
<td>5.17</td>
<td>1.74</td>
<td>2.03-74.3</td>
<td>0.80</td>
</tr>
<tr>
<td>All TE events</td>
<td>1.29</td>
<td>4.86</td>
<td>16.69</td>
<td>2.20-127</td>
<td>0.007</td>
</tr>
<tr>
<td>Thrombosis</td>
<td>0.00</td>
<td>1.39</td>
<td>2.86</td>
<td>1.37-5.98</td>
<td>0.005</td>
</tr>
<tr>
<td>Major bleed, TE event, thrombosis</td>
<td>2.91</td>
<td>8.33</td>
<td>2.86</td>
<td>1.37-5.98</td>
<td>0.005</td>
</tr>
<tr>
<td>Sudden death</td>
<td>0.29</td>
<td>0.35</td>
<td>1.19</td>
<td>0.07-19.06</td>
<td>0.90</td>
</tr>
<tr>
<td>Valve-related mortality</td>
<td>0.87</td>
<td>0.69</td>
<td>0.79</td>
<td>0.13-4.76</td>
<td>0.80</td>
</tr>
<tr>
<td>Total mortality</td>
<td>1.16</td>
<td>1.74</td>
<td>1.49</td>
<td>0.40-5.55</td>
<td>0.60</td>
</tr>
</tbody>
</table>
Results (stand-alone)

• Dual anti-platelet group (Groups 1-8)
  • How many people were in this group? 99
  • How many major bleeding and neurological events did they experience? 15
  • What is the “event rate” for them? 15%

• How many were in warfarin group? (Groups 9-16)
  • How many people were in this group? 102
  • How many major bleeding neurological events did they experience? 10
  • What is the event rate for them? 10%
Results (comparative)

• A (Groups 1-8): Consider warfarin + aspirin as control, DAPT as intervention
  • What is ARR
  • What is RR
  • What is RRR
  • What is NNT

• B (Groups 9-16): Consider DAPT as control, warfarin + aspirin as intervention
  • What is ARR
  • What is RR
  • What is RRR
  • What is NNT
Results (comparative)

- Using warfarin + aspirin as a control, DAPT as the intervention (A: Groups 1-8):

<table>
<thead>
<tr>
<th></th>
<th>Warfarin + ASA</th>
<th>DAPT</th>
<th>ARR</th>
<th>RR</th>
<th>RRR</th>
<th>NNT†</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patients</td>
<td>102</td>
<td>99</td>
<td>-5%</td>
<td>150%</td>
<td>-50%</td>
<td>-20</td>
</tr>
<tr>
<td>Events</td>
<td>10</td>
<td>15</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rate*</td>
<td>10%</td>
<td>15%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Minor rounding was introduced

†A negative NNT is sometimes better thought of “number needed to harm” – not all treatment is beneficial
Results (comparative)

• Using DAPT as a control, warfarin + aspirin as the intervention (B: Groups 9-16):

<table>
<thead>
<tr>
<th>Warfarin</th>
<th>ARR</th>
<th>RR</th>
<th>RRR</th>
<th>NNT</th>
</tr>
</thead>
<tbody>
<tr>
<td>+ ASA</td>
<td>5%</td>
<td>67%</td>
<td>33%</td>
<td>20</td>
</tr>
<tr>
<td>Patients</td>
<td>102</td>
<td>99</td>
<td>15</td>
<td></td>
</tr>
<tr>
<td>Events</td>
<td>10</td>
<td>15</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rate*</td>
<td>10%</td>
<td>15%</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Minor rounding was introduced
Decision time!

“DO. OR DO NOT. THERE IS NO TRY.”

—Yoda
Didactic Points

• Tools I used on purpose
  • Use words, formulas, pictures (graphs) for the same concept
  • A real story (your own, a family member, a patient …)
  • Humor (although you have to be funny – I’m working on it…)

• You can outright ignore large parts of a paper to make a point
  • Corollary: You do not have to address every topic raised by every paper every time

• You can get your own numbers from what is reported, even if the authors did not calculate them for you

• Work out calculations in advance

• Attempts at interactivity (COVID makes this challenging)

• Papers come as they come; very few are “perfect”