GIM Grand Rounds

Evidence Based Medicine - Putting Principles into Action

Brandon P. Combs, MD
Division of General Internal Medicine
University of Colorado School of Medicine
May 2014
1. If $p > 0.05$ for an RCT comparing two interventions, the study is conclusively negative

2. Absolute risk reduction is mathematically the same as “chance of benefit”

3. To calculate an NNT for a given intervention, all you need is starting risk and a corresponding RRR

4. Patients at high risk of a given outcome typically benefit more from treatment compared to those at lower risk even when RRR is the same

5. For very low prevalence conditions (< 1/1000), positive test results are generally false positives even when a test’s sensitivity and specificity are 99%
1. Review rationale for evidence based medicine (EBM)
2. Consider where to find the **BEST** evidence
3. Get comfortable with common biostats
4. **PRACTICE!!**
With EBM, there’s no turning back!
What is EBM?

• Integration of **clinical expertise, scientific evidence, and patient preferences** in order to make optimal decisions for patients
• Values **quantitative** data over anecdote or expert opinion
• A hierarchy of evidence
Hierarchy of Evidence

- Systematic reviews of RCTs
- Single randomized trial
- Observational studies
- Physiologic studies (LDL, FEV1, BMD, etc)
RCT

“Does it work?”

OBSERVATIONAL

“Is it dangerous?”
<table>
<thead>
<tr>
<th>Treatment</th>
<th>Mean Baseline</th>
<th>Adjusted Mean Change in A1C (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Onglyza 2.5 mg</td>
<td>8.4%</td>
<td>-0.9%</td>
</tr>
<tr>
<td>Placebo</td>
<td>8.1%</td>
<td>-0.4%</td>
</tr>
</tbody>
</table>

**Greater reduction with Onglyza 2.5 mg**

(95% CI: -0.7, -0.1)
Evidence alone is *never* enough!

- Are antibiotics effective for treatment of pneumococcal pneumonia?
- Are vitamin K antagonists effective for prevention of stroke in atrial fibrillation?
- Can you think of a situation where you might avoid these effective treatments?
1. Have confidence we’re providing best care for our patients

   Better evidence \(\rightarrow\) Greater confidence

2. Avoid unnecessary care/waste

3. Promote skepticism – eg niacin to raise HDL, ezetimibe to lower LDL (common practices \(\rightarrow\) RCTs demonstrate no benefit)
Finding the best evidence “4S” – *highly evolved clinicians need highly evolved sources*

INDIVIDUAL STUDIES (Pubmed) ARE LAST!!
Critical Appraisal — because sometimes you just have to

- Sometimes we have to assess the evidence ourselves
- Takes time
- LAST resort (if it hasn’t been done for us already)
Critical! Appraisal

• Is it **VALID**? (JAMA User’s Guides)

• What are the **RESULTS**?

• Are the results **IMPORTANT**?
About those results...

For a *diagnostic test* – is it accurate?

Compared to what?

A few quick examples?

Cue the 2x2 table!
The 2X2 is your friend

Gold Standard

Figure adapted from H. Gilbert Welch, MD, MPH
Sensitivity & Specificity

= “How often is the test right?”

Figure adapted from H. Gilbert Welch, MD, MPH
In 1,000 patients low risk by Wells’...

**how often is the d-dimer right?**

<table>
<thead>
<tr>
<th></th>
<th>PE Present</th>
<th>PE Absent</th>
<th>Predictive Values</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>D-dimer pos</strong></td>
<td>95</td>
<td>288</td>
<td>PPV:</td>
</tr>
<tr>
<td><strong>D-dimer neg</strong></td>
<td>5</td>
<td>612</td>
<td>NPV:</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>100</td>
<td>900</td>
<td></td>
</tr>
</tbody>
</table>

**Sensitivity**:

NPV = TN/TN+FN = 99%

**Specificity**:

PPV = TP/TP+FP = 25%
In a first-of-its-kind study, researchers have developed a blood test for Alzheimer's disease that predicts with astonishing accuracy whether a healthy person will develop the disease.

"This is a potential game-changer," said Dr. Howard Federoff, senior author of the report

"My level of enthusiasm is very high."

If study cohort’s 5% rate of conversion from normal cognition to MCI or Alzheimer’s disease is representative..., then test would have positive predictive value of just 35%. That is, nearly two-thirds of positive screening results would be false.
The probability that a given test result would be expected in a patient with disease compared to the probability that the same result would be expected in a patient without disease

Think of some tests with really high or low likelihood ratios
Likelihood ratios are easily calculated!

\[ LR^+ = \frac{\text{sensitivity}}{1-\text{specificity}} \]

\[ LR^- = \frac{1-\text{sensitivity}}{\text{specificity}} \]
**Rules of thumb:**

LR > 1 = disease more likely
LR < 1 = disease less likely

LR < 0.1 - rule out (neg d-dimer)
LR > 10 - rule in (PE seen on CT)

**How much more/less likely?**

LR 2 → 5 → 10
increases post test prob by 15% → 30% → 45%

LR 1/2 → 1/5 → 1/10
decreases post test prob by 15% → 30% → 45%
D-dimer
LR - = 0.07
LR + = 2.97

CTPE
LR+ = 20

Graph showing the relationship between pre-test probability, likelihood ratio, and post-test probability. The graph is labeled with the term 'PE' at the point of intersection.
About those results...

• For a *therapy trial* – what was magnitude of effect?
• Compared to what?
• Common ways to express *effect size*
A drug reduces risk of dying from heart disease from 20% to 15% over 10 yrs compared to placebo.

Which is true?

A. Drug reduces risk by 25%
B. Drug reduces risk by 5 percentage points
C. 5 out of 100 people taking drug will avoid death from heart disease (NNT 20)
D. Risk as a result of drug is 75% starting risk
E. All are true
“Stop Attack” for MI over 1 yr

Percentage of Heart Attacks That Result in Death

RRR: 67%

ARR: 50%

NNT: 2

Copyright © 2008 by David Newman
Aspirin reduces CV events in high risk patients

RRR: 20%
ARR: 2.5%
NNT: 40

Percentage Who Suffer a Heart Attack, Stroke, or Death

87 percent do not have a heart attack or stroke or die, regardless of therapy (unaffected)

2.5 percent taking aspirin avoid a heart attack or stroke or death (affected)

10.5 percent had a heart attack or stroke or died, regardless of therapy (unaffected)
Is this a good deal?
In patients with multiple risk factors for heart disease, 

**Lipitor** reduces risk of heart attack by **36%***

If you have risk factors such as family history, high blood pressure, age, low HDL (good cholesterol) or smoking.

That means in a large clinical study, 3% of patients taking a sugar pill or placebo had a heart attack compared to 2% of patients taking Lipitor.
Compare drugs A, B, and C below. Which drug is best, assuming equal side effects?

<table>
<thead>
<tr>
<th></th>
<th>Control Deaths</th>
<th>Treatment Deaths</th>
<th>RRR</th>
<th>ARR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Drug A</td>
<td>10%</td>
<td>8%</td>
<td>0.2 (20%)</td>
<td>.02 (2%)</td>
</tr>
<tr>
<td>Drug B</td>
<td>20%</td>
<td>16%</td>
<td>0.2 (20%)</td>
<td>.04 (4%)</td>
</tr>
<tr>
<td>Drug C</td>
<td>50%</td>
<td>40%</td>
<td>0.2 (20%)</td>
<td>.10 (10%)</td>
</tr>
</tbody>
</table>
Building confidence with confidence intervals
Point vs interval estimates

• Estimate the average height of colleagues here today – level of confidence?
• Can you suggest a range of possibilities? How confident are you now? 95%?
• The interval estimate is a more realistic view of the results AND incorporates statistical significance!
CI’s sometimes reported like this…

“Trends” indicate insufficient power
What does the p-value mean when reported with point estimate of a treatment effect? What about with a CI?

\[ p = \text{the probability that observed differences are due to chance} \]

\[ p < 0.05 \text{ implies the 95\% CI of the RR does not contain the value 1.0} \]
What’s the advantage of reporting treatment effects as interval estimates instead of point estimates with p-values?

95% CIs are intuitive and reveal an estimated treatment effect while p-values convey no clinical information.
Using confidence intervals

- Definitively positive, $p < 0.05$
- Definitively negative, $p > 0.05$
- Inconclusive, $p < 0.05$
- Inconclusive, $p > 0.05$

Risk Difference (ARR Death)


Copyright © American Medical Association. All rights reserved.
### Alendronate 10 mg/day for 1-4 years - Summary of Findings for Primary Prevention

<table>
<thead>
<tr>
<th>Outcomes</th>
<th>Illustrative comparative risks* (95% CI)</th>
<th>Relative effect (95% CI)</th>
<th>Number of participants (studies)</th>
<th>Quality of the evidence (GRADE)</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Assumed risk</td>
<td>Corresponding risk</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Without alendronate</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Vertebral fractures</strong></td>
<td>Low-risk population</td>
<td></td>
<td>RR 0.55 (0.38 to 0.80)</td>
<td>4576 (2 studies)</td>
<td>(++)++ Moderate</td>
</tr>
<tr>
<td>12 per 1000</td>
<td>7 per 1000 (5 to 10)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Moderate-risk population</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>53 per 1000</td>
<td>29 per 1000 (20 to 42)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Hip fractures</strong></td>
<td>Low-risk population</td>
<td></td>
<td>RR 0.79 (0.44 to 1.44)</td>
<td>4576 (2 studies)</td>
<td>(++++) Low</td>
</tr>
<tr>
<td>4 per 1000</td>
<td>3 per 1000 (2 to 6) Not statistically significant</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Moderate-risk population</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19 per 1000</td>
<td>15 per 1000 (8 to 27) Not statistically significant</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Outcomes</th>
<th>Illustrative comparative risks* (95% CI)</th>
<th>Relative effect (95% CI)</th>
<th>Number of participants (studies)</th>
<th>Quality of the evidence (GRADE)</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Assumed risk</td>
<td>Corresponding risk</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Without alendronate</td>
<td>With alendronate</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vertebral fractures</td>
<td>Moderate-risk population</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>53 per 1000</td>
<td>29 per 1000 (23 to 37)</td>
<td>RR 0.55 (0.43 to 0.69)</td>
<td>2785 (4 studies)</td>
<td>++ + + +</td>
</tr>
<tr>
<td></td>
<td>112 per 1000</td>
<td>62 per 100 (48 to 77)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High-risk population</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hip fractures</td>
<td>Moderate risk population</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>19 per 1000</td>
<td>19 per 1000 (5 to 16)</td>
<td>RR 0.47 (0.26 to 0.85)</td>
<td>5376 (5 studies)</td>
<td>++ + + +</td>
</tr>
<tr>
<td></td>
<td>87 per 1000</td>
<td>41 per 1000 (23 to 74)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

And Finally... *the art of medicine*

- Does the evidence you have reviewed MATTER?
- Were the endpoints IMPORTANT?
- Could it change how you care for your patient?
1. If $p > 0.05$ for an RCT comparing two interventions, the study is conclusively negative.

2. Absolute risk reduction is mathematically the same as “chance of benefit”.

3. To calculate an NNT for a given intervention, all you need is starting risk and a corresponding RRR.

4. Patients at high risk of a given outcome typically benefit more from treatment compared to those at lower risk even when RRR is the same.

5. For very low prevalence conditions ($< 1/1000$), positive test results are generally false positives even when a test’s sensitivity and specificity are 99%.
If a test to detect a disease whose prevalence is 1/1000 has a false positive rate of 5%, what is the chance that a person found to have a positive result actually has the disease, assuming you know nothing about the person's symptoms or signs? Assume test sensitivity is 100%.
If a test to detect a disease whose prevalence is 1/1000 has a false positive rate of 5%, what is the chance that a person found to have a positive result actually has the disease, assuming you know nothing about the person's symptoms or signs? Assume test sensitivity is 100%.

<table>
<thead>
<tr>
<th></th>
<th>Disease present</th>
<th>Disease absent</th>
<th>Predictive Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test pos</td>
<td>1</td>
<td>49</td>
<td>PPV = 2%</td>
</tr>
<tr>
<td>Test neg</td>
<td>0</td>
<td>950</td>
<td>NPV = 100%</td>
</tr>
<tr>
<td>Total</td>
<td>1</td>
<td>999</td>
<td></td>
</tr>
</tbody>
</table>
Thanks, everybody!