Intro to EBM

Part II: Critical Appraisal of the Literature

David J. Theige, M.D.

Objectives

- How to select an information resource
- Critical appraisal: diagnostic tests
- Biostatistical concepts about diagnostic tests
- Critical appraisal: therapy
- Biostatistical concepts about therapy

What is EBM?

- A systematic approach to clinical problem solving which allows the integration of the best available research evidence with clinical expertise and patient values.

Usefulness Equation

Usefulness of medical information = (Relevance * Validity) / Work


Factors in selecting an information resource

- Time limitations
- Depth of information needed
- Accessibility of information resources
- Skill in using information resources
From Evidence to Evidence-Based Resources

Review Articles

• Most are unsystematic—the author doesn’t ask a focused question and/or doesn’t examine all of the evidence (1st literature)

Systematic Review

• a subset of review articles
• formal approach to gathering, interpreting, and presenting the evidence
• methodologic standards enhance the likelihood that they will produce valid conclusions based on the best available evidence

CMAJ. 1988;138:697-703.

Types of Review Articles

Clinical Practice Guidelines

• systematically developed statements to assist practitioner and patient decisions about appropriate health care for specific clinical circumstances

Critical Appraisal

1) Is the evidence valid?
   - Are the study methods strong enough so that the results are likely to be true?

2) Is the evidence relevant?
   - Are the results important/clinically significant?
   - Are the results applicable to your patient population?
Diagnostic Tests

Critical Appraisal of the Literature

Is the evidence about this diagnostic test valid?

- Was there an independent, blind comparison with a reference ("gold") standard of diagnosis?
- Was the test evaluated in an appropriate spectrum of patients (like those in whom we would use it in practice)?
- Was the reference standard applied regardless of the test result?

Is the evidence about this test clinically significant?

- Does this (valid) evidence demonstrate an important ability of this test to accurately distinguish patients who do and don’t have a specific disorder?
  - Sensitivity
  - Specificity
  - Likelihood Ratio

2 x 2 Table

<table>
<thead>
<tr>
<th></th>
<th>Target Disorder Present</th>
<th>Target Disorder Absent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test Result</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Positive</td>
<td>a (true positive)</td>
<td>b (false positive)</td>
</tr>
<tr>
<td>Negative</td>
<td>c (false negative)</td>
<td>d (true negative)</td>
</tr>
</tbody>
</table>

Sensitivity

- The proportion of patients with the target disorder who have a positive test result.
- PID ("Positive in Disease")
- \( \frac{a}{a + c} \)

Specificity

- The proportion of patients without the target disorder ("healthy") who have a negative test result.
- NIH ("Negative in Health")
- \( \frac{d}{b + d} \)
Pretest Probability (Prevalence)

- The proportion of patients who have the target disorder before the test is carried out.
- \( \frac{(a + c)}{(a + b + c + d)} \)
- Often requires the clinician to estimate.

Positive Predictive Value

- \( \frac{a}{(a + b)} \)
- The probability that a person known to have a positive test result does have the target disorder.

Negative Predictive Value

- \( \frac{d}{(c + d)} \)
- The probability that a person known to have a negative test result does NOT have the target disorder.

Can I apply this valid, important test to a specific patient?

- Is the diagnostic test available, affordable, and accurate in our setting?
- Can we generate a clinically sensible estimate of our patient’s pretest probability?
- Will the resulting post-test probabilities affect our management and help our patient?

Relevance

- Based on the prevalence of a problem, the type of outcome, and the magnitude and precision of the results.
- Outcomes may be:
  - Patient-oriented (survival, symptoms, quality of life, cost)
  - Or disease-oriented (laboratory measurements or surrogate end-points)
Are the results of this therapeutic trial valid?

- Was the assignment of patients to treatment randomized?
- Was follow-up of patients sufficiently long and complete?
- Were all patients analyzed in the groups to which they were randomized?

Are the results of this therapeutic trial valid? (2\textsuperscript{nd} Criteria)

- Were patients and clinicians kept blind to treatment?
- Were groups treated equally, apart from the experimental treatment?
- Were the groups similar at the start of the trial?

What are the results of this therapeutic trial?

- (Population? Treatment and duration?)
- What is the magnitude of the treatment effect?
- How precise is the estimate of the treatment effect?
- (To achieve what outcome?)
- (Duration of follow-up before outcomes measured?)

What is the magnitude of the treatment effect?

- Control Event Rate
- Experimental Event Rate
- Relative risk reduction
  \[ \text{RRR} = \frac{(\text{CER} - \text{EER})}{\text{CER}} \]
- Absolute risk reduction
  \[ \text{ARR} = \text{CER} - \text{EER} \]
- Number Needed to Treat
  \[ \text{NNT} = \frac{1}{\text{ARR}} \]

Terms Relating to Rx Effects

- When the Experimental Rx
  - reduces the risk for a bad event: RRR, ARR, NNT
  - increases the probability of a good event:
    \[ \text{Relative Benefit Increase (RBI)} \]
    \[ \text{Absolute Benefit Increase (ABI)} \]
    \[ \text{NNT} \]
  - increases the probability of a bad event:
    \[ \text{Relative Risk Increase (RRI)} \]
    \[ \text{Absolute Risk Increase (ARI)} \]
    \[ \text{Number Needed to Harm (NNH)} \]

Are the results of this therapeutic trial clinically significant?

- Compare NNT to other treatments (and durations)
- Consider dimension of follow-up time
- Temper with clinical experience and expertise
- Adverse effects can be handled similarly by using NNH
How Precise is the Estimate of the Treatment Effect?

- NNTs are estimates and we should specify limits within which the true NNT lies.
- Confidence Interval (CI)
  - The true NNT lies within the 95% CI 95% of the time.
  - The smaller the number of patients in the trial, the wider the confidence interval.

Confidence Intervals

- An estimate of "sampling variation"
  - The idea that the same study carried out on different samples of patients would not yield the same results, but their results would be spread around the true, but unknown value.
- CI gives a measure of precision
- CI quantifies both the effect of interest, and the degree of uncertainty about the effect
- 95% CI commonly used by convention

P values

- Measure the strength of the evidence against the null hypothesis ("no effect").
- Tell us nothing about the magnitude or direction of an effect (statistical significance vs. clinical significance)

In contrast to P values

- Confidence Intervals quantify both:
  - the effect of interest
    - Sensitivity and specificity of a diagnostic test
    - NNT of a treatment
  - the degree of uncertainty in this effect

Are the valid, important results applicable to our patient?

- Is our patient so different from those in the study that its results don’t apply?
- Is the treatment feasible in our setting?
- What are our patient’s potential benefits and harms from the therapy?
- What are our patient’s values and expectations for both the outcome we are trying to prevent and the treatment we are offering?

References

- JAMA evidence https://jamaevidence-mhmedical-com.esproxy.library.und.edu/
- Centre for Evidence-Based Medicine. http://www.cebm.net/critical-appraisal/