Critical Appraisal

Patarawan Woratanarat, MD, PhD (Clin.Epid.)
Department of Orthopaedics
Faculty of Medicine Ramathibodi Hospital
Mahidol University

Critical Appraisal

EBM group

Users' Guides to the Medical Literature

A second edition of the Users' Guides to the Medical Literature is in the way. The authors, a crew similar to those who produced the first Users' Guides, with the same format, have been asked to produce this book for the Evidence-Based Medicine Working Group. All in the group are expected to provide a manuscript by mid-2000.

What's New in the Second Edition?

New chapters based on recent EBM and JAMA publications from the CLARITY group will help readers to:

- Understand innovations in design of non-pharmacologic randomized trials;
- Navigate the diverse world of clinical trial design, and;
- Learn how to interpret tools that rely on computer databases, and

Avoid being misled by biased presentations of research findings.

An additional new chapter will address how to change clinical behavior to improve patient care. Readers will gain a sophisticated understanding of the latest methodological advances in conducting studies and interpreting their results.

If you would like to provide feedback or comments on the new edition, send your comments to Dr. Gordon Guyatt:
guyatt@mcmaster.ca

Choosing Evidence Worksheet

I1) Diagnostic Tests

Based on the Users' Guides to the Medical Literature: A Manual for Evidence-Based Clinical Practice, this worksheet can serve as an aid to the critical appraisal of an article about interpreting diagnostic test results.

Are the results valid?

☑️ ✗ Did clinicians face diagnostic uncertainty?

Are the results valid?

☑️ ✗ Was there a blind comparison with an independent gold standard applied similarly to the treatment group and the control group?
Critical appraisal

- Treatment → RCT
- Prognosis → cohort study
- Risk/harm → cohort/case-control study
- Diagnostic test → cross sectional study

- Systematic review
- Decision analysis
- Economic analysis

3 Steps of Critical Appraisal

- Valid (close to the truth)
  Rating of validity → if not → stop
- Results
- Application

Treatment

- Are the results valid?
  - Did experimental and control groups begin the study with a similar prognosis?
    - Were patients randomized?
    - Was randomization concealed?
      - opaque envelop, internet, call research center
    - Were patients analyzed in the group to which they were randomized?
    - Treatment & control groups have similar prognosis?

Treatment

- Are the results valid?
  - Did experimental and control groups retain a similar prognosis after the study get started?
    - Were patients aware of group allocation? –blind?
    - Was clinicians aware of group allocation? –blind?
    - Were outcome assessors aware of group allocation? –blind?
  - Was the follow-up complete?
  Poor     Good
Treatment

- What are the results?
  - How large was the treatment effect?
    - Relative risk reduction
    - Absolute risk reduction
  - How precise?
    - 95% confidence interval
    - P-value

- How can I apply the results to patient care?
  - Were the study patients similar to the patient in my practice?
  - Were all clinically important outcomes considered?
  - Are likely treatment benefits worth the potential harm and costs?
    - NNT, NNH
    - Cost

NNT NNH

- Number needed to treat (NNT)
  - the average number of patients that a doctor would need to treat in order to have one additional event occur
- Number needed to harm (NNH)
  - an increase in bad events like side effects that might be associated with a treatment

Example

- Pediatric femoral shaft fracture
  - Hip spica group had malunion 40%
  - External fixator group had malunion 15%
    - Absolute risk reduction (ARR) = 0.40-0.15 = 0.25
    - NNT = 1/ARR = 1/0.25 = 4
  - Hip spica group had pin tract infection 0%
  - External fixator group had pin tract infection 40%
    - Absolute risk reduction (ARR) = 0.40-0 = 0.40
    - NNH = 1/ARR = 1/0.40 = 2.5
Exercise

Group discussion

Select one!
- ข้อเข่าเสื่อม **glucosamine sulfate**
- ความจำเสื่อม **Ginko biloba**
- เป็นหวัด **Vitamin C**

Prognosis
- Are the results valid?
  - Was the sample of patients representative?
  - Were the patients sufficiently homogeneous with respect to prognostic risk?
  - Was the follow-up complete?
  - Were the objective and unbiased outcome criteria used?

Prognosis
- What are the results?
  - How likely are the outcomes over time?
    - Survival analysis, hazard ratio
  - How precise of the estimates of likelihood?
    - 95% confidence interval
Prognosis

How can I apply the results to patient care?
- Were the study patients and their management similar to those in my practice?
- Was follow-up sufficiently long?
- Can I use the results in the management of patients in my practice?

Cohort study

<table>
<thead>
<tr>
<th>Exposure</th>
<th>Disease</th>
<th>No disease</th>
<th>Total</th>
<th>No. of cases</th>
<th>Poor work</th>
<th>Good work</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>+</td>
<td>a</td>
<td>b</td>
<td>a+b</td>
<td>&gt; 10</td>
<td>80</td>
<td>10</td>
<td>90</td>
</tr>
<tr>
<td>-</td>
<td>c</td>
<td>d</td>
<td>c+d</td>
<td>&lt; 10</td>
<td>20</td>
<td>90</td>
<td>110</td>
</tr>
<tr>
<td>a+c</td>
<td>b+d</td>
<td>n</td>
<td>100</td>
<td>100</td>
<td>200</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Term</th>
<th>General</th>
<th>Example</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Attributable risk</td>
<td>a/(a+b) - c/(c+d)</td>
<td>80/90 - 20/110 = 0.7</td>
<td>The incidence of disease attributable to exposure</td>
</tr>
<tr>
<td>Relative risk</td>
<td>a/(a+b) + c/(c+d)</td>
<td>80/90 + 20/110 = 5</td>
<td>Exposed persons are more likely to become diseased, relative to non exposed persons</td>
</tr>
</tbody>
</table>

Statistical analysis

<table>
<thead>
<tr>
<th>จับคู่ตัวอย่าง</th>
<th>ข้อมูลคัดเก็บในปัจจุบัน</th>
<th>ข้อมูลคัดเก็บในปัจจุบัน</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 กลุ่ม</td>
<td>Mean ± Standard deviation</td>
<td>Proportion, percentage</td>
</tr>
<tr>
<td>2 กลุ่ม</td>
<td></td>
<td></td>
</tr>
<tr>
<td>- ข้อมูลเป็นอิสระต่อกัน (independent)</td>
<td>Unpaired T-test</td>
<td>Chi-square</td>
</tr>
<tr>
<td>- ข้อมูลมีการจับคู่กลุ่มประชากร (matched pair)</td>
<td>Paired T-test</td>
<td>McNemar's Chi-square</td>
</tr>
<tr>
<td>หรือ เป็นการเก็บข้อมูลเรียบเรียงกันและหลักการทั่วไป</td>
<td>Analysis of variance</td>
<td>Chi-square</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>จับคู่ตัวอย่าง</th>
<th>ข้อมูลคัดเก็บในปัจจุบัน</th>
<th>ข้อมูลคัดเก็บในปัจจุบัน</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 กลุ่ม</td>
<td>Sign test</td>
<td>Proportion, percentage</td>
</tr>
<tr>
<td>2 กลุ่ม</td>
<td></td>
<td></td>
</tr>
<tr>
<td>- ข้อมูลเป็นอิสระต่อกัน (independent)</td>
<td>Mann-Whitney U test</td>
<td>Fisher's exact</td>
</tr>
<tr>
<td>- ข้อมูลมีการจับคู่กลุ่มประชากร (matched pair)</td>
<td>Wilcoxon sign-rank test</td>
<td>McNemar's test</td>
</tr>
<tr>
<td>หรือ เป็นการเก็บข้อมูลเรียบเรียงกันและหลักการทั่วไป</td>
<td>Kruskall-Wallis</td>
<td>Fisher's exact</td>
</tr>
</tbody>
</table>
Diagnostic test

- Are the results valid?
  - Did clinicians face diagnostic uncertainty?
    - Were subjects drawn from a group in which it is not known whether the condition of interest is present/absent?

- Was there a blind comparison with an independent gold standard applied similarly to the treatment and control group?
  - Acceptable gold standard?
  - Independent test and gold standard?
  - Assess without knowledge of the other results?

Diagnosis

Rating the validity

- Poor
- Good
Diagnostic test

- What are the results?
  - What range of likelihood ratios were associated with the range of possible test results?

Likelihood Ratio is the likelihood of a given test result in a patient with the target disorder compared to the likelihood of the same result in a patient without that disorder.

Likelihood ratio

- Example
  - A man with tibial fracture → open fx???
    - Bony exposure → ? %
    - Fat globule → ? %
    - Wound 10 cm → ? %
    - Pulseless → ? %

The likelihood ratio

- The likelihood ratio incorporates both the sensitivity and specificity of the test and provides a direct estimate of how much a test result will change the odds of having a disease.
- The likelihood ratio for a positive result (LR+) tells you how much the odds of the disease increase when a test is positive.
- The likelihood ratio for a negative result (LR-) tells you how much the odds of the disease decrease when a test is negative.

The likelihood ratio

- The likelihood ratio of a positive test result (LR+) is sensitivity divided by 1 - specificity.
- The likelihood ratio of a negative test result (LR-) is 1 - sensitivity divided by specificity.
Diagnostic test

<table>
<thead>
<tr>
<th>Test</th>
<th>Disease</th>
<th>No</th>
<th>Total</th>
<th>SLN</th>
<th>ALN</th>
<th>No</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>+</td>
<td>a</td>
<td>b</td>
<td>a+b</td>
<td>+</td>
<td></td>
<td>58</td>
<td>0</td>
</tr>
<tr>
<td>-</td>
<td>c</td>
<td>d</td>
<td>c+d</td>
<td>-</td>
<td>11</td>
<td>138</td>
<td>149</td>
</tr>
<tr>
<td></td>
<td>a+c</td>
<td>b+d</td>
<td>n</td>
<td></td>
<td>69</td>
<td>138</td>
<td>207</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Term</th>
<th>General</th>
<th>Example</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sensitivity</td>
<td>a/(a+c)</td>
<td>58/69 (84.1%)</td>
<td>Proportion of those with the condition who have a positive test</td>
</tr>
<tr>
<td>Specificity</td>
<td>B/(b+d)</td>
<td>138/138 (100%)</td>
<td>Proportion of those without the condition who have a negative test</td>
</tr>
<tr>
<td>Positive predictive value</td>
<td>a/(a+b)</td>
<td>58/58 (100%)</td>
<td>Proportion of those with a positive test who have the condition</td>
</tr>
<tr>
<td>Negative predictive value</td>
<td>d/(c+d)</td>
<td>138/149 (92.6%)</td>
<td>Proportion of those with a negative test who do not have the condition</td>
</tr>
<tr>
<td>Accuracy</td>
<td>a+d/n</td>
<td>196/207 (94.7%)</td>
<td>Proportion of accurate diagnostic test</td>
</tr>
</tbody>
</table>

The likelihood ratio

- Once you have specified the pre-test odds, you multiply them by the likelihood ratio. This gives you the post-test odds.
- \[ \text{Odds (post)} = \text{Odds (pre)} \times \text{LR} \]
Fagan Normogram

Chance of having ALN metastasis Before biopsy

Results of SLN pathonegative

7% chance of ALN metastasis

No further axillary Dissection & Rx as no ALN metas

Diagnosis

How can I apply the results to patient care?

- Will the reproducibility of the test results and its interpretation be satisfactory in your clinical setting?

Diagnosis

- How can I apply the results to patient care?
  - Are the results applicable to the patient in your practice?
    - Different test for different severity?
    - Different test for mixed condition?

Diagnosis

- How can I apply the results to patient care?
  - Will the results change your management strategy?
    - What are the test and treatment thresholds for the health condition to be detected?
    - LR high/low enough
Diagnosis

- How can I apply the results to patient care?
  - Will patients be better off as a result of the test?
  - Care differ for different test results?
  - The anticipated changes do good > harm?

Systematic review

- Are the results valid?
  - Did the review explicitly address a sensible clinical question?
  - Was the search for relevant studies detailed and exhaustive?
  - Were the primary studies of high methodologic quality?
  - Were assessment of studies reproducible?

Systematic review

- What are the results?
  - Were the results similar from study to study?
  - What are the overall results of the review?
  - How precise were the results?

Systematic review

- How can I apply the results to patient care?
  - How can I best interpret the results to apply them to the care of patients in my practice?
  - Were all clinically important outcome considered?
  - Are the benefit worth the costs and potential risks?
Risk/harm

Are the results valid?
- Did the investigators demonstrate similarity in all known determinants of outcome: did they adjust for differences in the analysis?
- Were exposed patients equally likely to be identified in the two groups?
- Were the outcomes measured in the same way in the groups being compared?
- Was follow-up sufficiently complete?

Poor          Good

Risk/harm

How can I apply the results to patient care?
- Were the study patients similar to the patient in my practice?
- Was the duration of follow-up adequate?
- What was the magnitude of the risk?
- Should I attempt to stop the exposure?

Case control study

<table>
<thead>
<tr>
<th>Exposure</th>
<th>Disease</th>
<th>No disease</th>
<th>Total</th>
<th>No. of cases</th>
<th>Poor work</th>
<th>Good work</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>+</td>
<td>a</td>
<td>b</td>
<td>a+b</td>
<td>&gt; 10</td>
<td>80</td>
<td>10</td>
<td>90</td>
</tr>
<tr>
<td>-</td>
<td>c</td>
<td>d</td>
<td>c+d</td>
<td>&lt; 10</td>
<td>20</td>
<td>90</td>
<td>110</td>
</tr>
<tr>
<td></td>
<td>a+c</td>
<td>b+d</td>
<td>n</td>
<td></td>
<td>100</td>
<td>100</td>
<td>200</td>
</tr>
<tr>
<td>Term</td>
<td>General</td>
<td>Example</td>
<td>Definition</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Odds ratio</td>
<td>ad/bc</td>
<td>80x90/20x10 = 36</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Summary

Critical appraisal has 3 steps
- Validity: see the methodology
- Results: see strength of the associations
- Application: see our population
Questions?