EBP STEP 2

APPRAISING THE EVIDENCE : So how do I know that this article is any good? (Quantitative Articles)

Alison Hoens
Clinical Assistant Prof, UBC
Clinical Coordinator, PHC

Maggie McIlwaine
Clinical Instructor, UBC
Clinical Coordinator, PT, BC Children’s Hospital
EBP - THE PROCESS

Clinical Problem

Ask

Acquire

Appraise

Act

Apply
EBP - WHAT IS IT?

- Evidence-Based Practice:
  - “The conscientious, explicit, and judicious use of current best evidence in making decisions about the care of individual pts”
  - “The integration of best research evidence with clinical expertise and patient values”
    - Sackett, 2000 Evidence-Based Medicine. How to Practice and Teach EBM. 2nd Ed. Churchill Livingstone
BARRIERS TO EBP


- **Ranking of importance of factors influencing current practice:**
  - Experience
  - Continuing education (practical)
  - Colleague Influence
  - Continuing Education (theory)
  - Professional Literature * secondary sources
  - Entry Level Training
BARRIERS TO EBP

- Lack of time, computing resources, not enough evidence, lack of access; lack of skills for searching, appraising, and interpreting; lack of incentives (Bennett S. et al, 2003. Australian OT Journal, 50, 13-22.)
- Publication bias, indexing issues, language issues, assessing internal validity, access to electronic databases, access to full text, assessing applicability, drawing conclusions (Maher. C. et al. Phys Ther, 84: 645-654).
ASK

- Foreground Questions
  - PICO
  - P - Patient/Problem
  - I - Intervention
  - C - Comparison
  - O - Outcome
Building the search strategy

- Identify concepts from PICO
- Decide on words
  - 2 methods: keywords or classification
  - Boolean operators
  - Truncation
  - Limits: study design; age; gender; language; year of publication
EVALUATING QUANTITATIVE ARTICLES

- Critical Review Form, Quantitative Studies, developed by McMaster OT Evidence-Based Practice Research Group (Law et al, 1998)
- User’s Guide To The Medical Literature. JAMA. (Guyatt & Rennie, 2002).
- BMJ (1997) series July-Sept (Greenhalgh, T)
- A checklist to evaluate a report of a non-pharmacological trial ... (Boutron et al, 2005)
EVALUATING QUANTITATIVE ARTICLES

- PEDro
Purpose

- Determine the purpose of the study

- Study Purpose
  - Clearly stated?
  - Phrased as a research question or hypothesis
Review of Literature

- Literature Review
  - Provides a synthesis of appropriate previous research & the clinical importance of the topic?
  - Includes primary > secondary sources
  - Interprets results of previous work
  - Clearly demonstrates the ‘holes’ which need to be filled by this particular study and thus justifies the need for this study
STUDY DESIGN

- **RCT**
  - Randomization to intervention & control grps
  - Prospective
  - Control (? ethical to withhold Rx)

- **Cohort**
  - Grp exposed to similar situation (program or diagnosis); observed over time *prospective
  - May have a comparison or control grp
  - Not random allocation; must find another grp of similar age, gender & other important factors
  - * caution re confounding factors
STUDY DESIGN

- **Single subject**
  - Single or multiple single subjects
  - Repeated measures before, during and after intervention
  - Subject serves as own control

- **Before-After**
  - Single subjects or groups
  - No control group
  - Useful when can’t withhold Rx
STUDY DESIGN

- **Case-control**
  - Retrospective
  - Explores what makes grps different
  - Compared to a comparable ‘control’ grp
  - Often problems with confounding variables

- **Cross-sectional**
  - One grp evaluated all at the same time
  - Useful when little is known
  - Eg. Surveys, questionnaires & interviews
  - Impossible to know if all factors included thus can’t draw cause-effect conclusions
Case-study
- Descriptive data about the relationship between an exposure and an outcome of interest
- No control grp
- Provides info for further studies

Appropriateness of Design
- Depends on:
  - Knowledge of topic/issue
  - Outcomes
  - Ethical issues
  - Study-purpose/question
LEVEL OF EVIDENCE

- Determine the level of evidence
LEVEL OF EVIDENCE

I a) Systematic review (with homogeneity) of RCTs
   b) individual RCT (with narrow confidence interval)
   c) all or none

II a) systematic review (with homogeneity) of cohort studies
    b) individual cohort study, including low quality RCTs eg. < 80% follow-up
    c) “outcomes” research

III a) systematic review (with homogeneity) of case-control studies
     b) Individual case-control studies

IV Case series (and poor-quality cohort & case-control studies)

V Expert opinion without critical appraisal or based on
    physiology, bench research or first principles’
     - Oxford Center for EBM levels of Medicine May 2001 *for Therapy
LEVEL OF EVIDENCE

Many hierarchies are available

1. Inconsistent nomenclature introduces a wide range of possibilities - aim of EBP is to reduce unnecessary inconsistencies!

2. The hierarchies give differential weight to consensus and evidence


Suggestion: use what makes sense - in a clinical setting: KISS
HIERARCHY OF EVIDENCE

- Meta-analysis
- Systematic reviews
- Narrative reviews
- Clinical Practice Guidelines
- Randomized Controlled Trials
- Cohort Studies
- Case-controlled Studies
- Case studies
- *Expert opinion
  - Modified from Cormack (2002)
LEVEL OF EVIDENCE

- **META ANALYSIS**
  - Locate clinical trials
  - Criteria: RCT
  - Rank: Methodological Score
  - Similar outcome measure
  - Variable Rx protocols
  - Pool results
  - Statistical analysis
  - ?Significant effect
    - Houghton, 2004

- **SYSTEMATIC REVIEWS**
  - Locate clinical trials
  - Criteria: RCT
  - Rank: Methodological score
  - Variable outcomes
  - Organized based on Rx protocols
  - Blinded reviewers (+/- or ?)
  - No statistics (#/+/−)
**LEVEL OF EVIDENCE**

- **NARRATIVE**
  - Question broad
  - Sources/search may not be specified; potentially biased
  - Selection of articles may not be specified; potentially biased
  - Appraisal variable
  - Synthesis often qualitative
  - Inferences maybe EB

- **SYSTEMATIC**
  - Question focused
  - Sources/search must be comprehensive and search strategy detailed
  - Selection of articles based on strict criteria
  - Appraisal rigorous
  - Quantitative summary (meta-analysis)
  - Inferences EB

*Table 7-1 (Law et al, 2002) p. 111*
STRENGTH OF EVIDENCE

- Quality of methodology
- "If you are deciding whether a paper is worth reading, you should do so on the design of the methods section and not on the interest of the hypothesis, the nature or impact of the results, or the speculation in the discussion"

Many scales for strength of evidence

* Boutron et al, (2005)
* Downs & Black (1998)
* Van der Heijden et al. (1996)
* Sindhu et al, (1997)
* De Vet et al, (1997)
* PEDro scale with explanations
  http://www.pedro.fhs.usyd.edu.au/scale_item.html
STRENGTH OF EVIDENCE

- Eg. Medlicott & Harris 2006:
  - 1. Randomization
  - 2. Inclusion & exclusion criteria
  - 3. Similarity of grps at baseline
  - 4. Protocol described sufficiently for replicability
  - 5. Reliability of data with outcome measures investigated
  - 6. Validity of data addressed
  - 7. Blinding of patient, Rx provider & assessor
  - 8. Long-term (6 m) results assessed
  - 9. Adherence to home prgm assessed
    - Yes vs No: Strong (8-10 Yes); Mod (6-7 Yes); Weak (<=5 Yes)
5.1 The Sample

- Was there a detailed description eg. Age, gender, duration of disease/disability? *mean, SD, range! 
- Were the number reported? Were the groups equal & similar? Should they be stratified? 
- Was there a description of how subjects were sampled/recruited?
5.1 The Sample Cont’d

- Were there appropriate inclusion/exclusion criteria?
- Was there justification of sample size?
- If there was more than 1 grp, were subjects randomized?
- Was allocation concealed?
- Were the ethics procedures reported?
5.1 Cont’d: Sample selection biases
- Volunteer or referral bias
- Seasonal bias
- Attention bias
INTERPRETING THE RESULTS OF CLINICAL TRIALS

- **Phase 1**: small numbers, assesses tolerance, safety issues
- **Phase 11**: estimates the efficacy, small scale, descriptive, pilot study
- **Phase 111**: large numbers, power size comparative trials
- **Phase 1V**: seeks to determine neg side effects.
ABSENCE OF EVIDENCE IS NO EVIDENCE OF ABSENCE
DANGER

TREATMENT A

TREATMENT B

OUTCOME VARIABLE FEV₁ SHOWS NO SIGNIFICANT DIFFERENCE

CONCLUSION:
TREATMENT A = TREATMENT B
TYPE 11 ERROR

- Infers two treatments are equal efficacy when they are not
- Need to do a power analysis prior to the study
- Usually need to increase the sample size to detect true differences
Negative Findings in Clinical Studies
Moher et al. JAMA 1994;272:122

Reviewed 383 studies
70 showed negative findings (no sig. Diff)
Only 36% (21 studies) had a Power size large enough to detect a true difference.
5.2 The Outcomes

- Were the outcomes clearly described?
- Were they detailed sufficiently for replication?
- Was the frequency of outcome measures described?
- Were the measures relevant to clinical outcome?
- Was reliability examined/reported & confirmed with these investigators?
- Was validity examined & reported (all relevant components - content validity; in relationship to gold standards - criterion validity)?
5.2 Cont’d: Measurement/detection biases

- No. of outcome measures - only 1 (eg. If excludes a critical measure) vs too many for the sample size
- Lack of ‘masked’/’blinded’ evaluation
- Recall or memory bias
SHORT TERM STUDIES

- Cannot imply long-term efficacy with short term studies
- Many outcome measures need a long duration to decrease confidence interval.
5.3 The Intervention

- Was the intervention described in detail (to replicate)?
- Was the intervention relevant?
- Who delivered it; were they trained? Were they blinded?
- Was the frequency appropriate?
- Was the setting appropriate?
- Was contamination/co-intervention avoided?
5.3 Cont’d: Intervention/Performance

Biases

- **Contamination** - some of the control group receives the intervention
- **Co-intervention** - intervention group receives other care that may impact outcome
- **Timing of intervention**
- **Site of treatment**
- **Different therapists**
Results

- Were the statistical methods provided in detail?
- Were the statistical methods appropriate?
- Were drop-outs reported?
RESULTS

- **Drop-outs:**
  - Important to determine what was done with their data
    - Excluded
    - Assumptions made for missing values
    - Analyzed separately
  - **Intention to treat analysis**
    - To reduce bias of when some pts drop out
    - All pts analyzed together as the Rx group whether or not they received the Rx or completed the study
RESULTS

- COMMON STATS TERMS
  - Descriptive statistics
    - Frequency distributions
    - Measures of central tendency
      - Mean - affected by extreme scores esp when total number of scores is small
      - Median - 50th percentile; less affected by extreme scores
      - Mode - most frequent score
    - Measures of variance/variability
      - Range - highest and lowest scores; affected by extremes
      - Variance - spread of scores around the mean
      - Standard deviation - square root of the variance
RESULTS

- Parametric vs nonparametric tests
  - Parametric (*for normal distributions)
    - T-test - compares means of two groups
    - Anova - compares means of groups in terms of variance
  - Nonparametric (*less powerful)
    - Wilcoxon signed rank test
    - Whitney-Mann-Wilcoxon test
    - Kruskal-Wallis test
RESULTS

**P value** - the probability that the observed diff btwn the 2 grps might have occurred by chance. Not an indication of accuracy.

**Confidence interval** - the range in which we are 95% certain that the true population Rx effect will lie. A measure of accuracy. Better than p value. If the CI does not cross zero it is statistically significant.

**Odds ratio** - a way of comparing whether the probability of a certain event is the same for two groups.

- 1 = the event is equally likely in both groups
- >1 = the event is more likely in the first group
- <1 = the event is less likely in the first group.
CONFIDENCE INTERVALS

Increase in Lifetime Mortality Risk
Effect of 10 μg/m³ PM₂.₅

% Increase

-10% 0% 5% 10% 15% 20% 25% 30%

Total Cardio-Pulmonary Lung Cancer All Other

Narrow or Wide
RESULTS

- **Relative Risk (RR)** - the number of pts who achieve the stated end point divided by the total no. of pts in the grp.
  - 1 = the event is equally likely in both groups
  - >1 = the event is more likely in the first group
  - <1 = the event is less likely in the first group.

- **Number needed to treat (NNT)** - the number of people who would need to be treated to prevent the event of interest. *smaller the better eg. 10 vs 100*
RESULTS

- **Fail-safe number** - an estimate of the number of unpublished studies that demonstrate the opposite finding in order to refute the findings of the current meta-analysis.

- **Regression equation** - statistical technique used to explain or predict the behavior of a dependent variable. $Y=a+bx+c$, Eg creating a mathematical equation to describe the ‘best fit’ of a line through data points that are not linear.

- **Forest Plot** - graph that visually shows the info from all the studies in the meta-analysis to indicate the overall effect of Rx.

- **Funnel plot** - statistical test to determine if there is likely to be ‘publication bias’.
## Impact of Treatment on Mortality

<table>
<thead>
<tr>
<th>Study name</th>
<th>Odds ratio</th>
<th>Lower limit</th>
<th>Upper limit</th>
<th>Odds ratio and 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kelly, 1964</td>
<td>0.590</td>
<td>0.096</td>
<td>3.634</td>
<td></td>
</tr>
<tr>
<td>Hedrin, 1980</td>
<td>0.464</td>
<td>0.201</td>
<td>1.074</td>
<td></td>
</tr>
<tr>
<td>Leigh, 1962</td>
<td>0.394</td>
<td>0.076</td>
<td>2.055</td>
<td></td>
</tr>
<tr>
<td>Novak, 1992</td>
<td>0.490</td>
<td>0.088</td>
<td>2.737</td>
<td></td>
</tr>
<tr>
<td>Saint, 1998</td>
<td>1.250</td>
<td>0.479</td>
<td>3.261</td>
<td></td>
</tr>
<tr>
<td>Pilbeam, 1936</td>
<td>0.129</td>
<td>0.027</td>
<td>0.605</td>
<td></td>
</tr>
<tr>
<td>Day, 1960</td>
<td>0.313</td>
<td>0.054</td>
<td>1.805</td>
<td></td>
</tr>
<tr>
<td>Kelly, 1966</td>
<td>0.429</td>
<td>0.070</td>
<td>2.620</td>
<td></td>
</tr>
<tr>
<td>Singh, 2000</td>
<td>0.718</td>
<td>0.237</td>
<td>2.179</td>
<td></td>
</tr>
<tr>
<td>Stewart, 1994</td>
<td>0.143</td>
<td>0.082</td>
<td>0.250</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.328</td>
<td>0.233</td>
<td>0.462</td>
<td></td>
</tr>
</tbody>
</table>

*Meta Analysis*
FUNNEL PLOT

Precision of estimate of treatment effect

Magnitude of the effect size

Precision of estimate of treatment effect

Favor Intervention  Favor Control

Outcome Measure
Conclusions

- Were clinical implications explored?
- Were conclusions restricted to a reasonable interpretation of the results?
- Were limitations of the study reported?
APPLICABILITY

- Ask yourself - how can I apply the results to patient care?
  - Were the study patients similar to my patient? (demographics, severity, co-morbidity)
  - Were all clinically important outcomes considered?
  - Are the likely treatment benefits worth the potential harm and costs?
USEFUL RESOURCES

- Center for Evidence-Based Medicine
  http://www.cebm.net

- PEDro scale with explanations
  http://www.pedro.fhs.usyd.edu.au/scale_item.html
FINAL SUGGESTION

- Start/join a journal club!