PREP #30:
Systematic reviews, meta-analysis and critical reading of medical literature:
Evidence-Based Medicine

Phyllis W. Speiser, MD
Chief, Div Ped Endo, CCMC
Professor of Pediatrics
Hofstra-NSLIJ School of Medicine
The North Shore LIJ Health System adheres to the ACCME’s new Standards for Commercial Support. Any individuals in a position to control the content of a CME activity, including faculty, planners, and managers, are required to disclose all financial relationships with commercial interests. All identified potential conflicts of interest are thoroughly vetted by the North Shore-LIJ for fair balance and scientific objectivity and to ensure appropriateness of patient care recommendations.

- Course Director and Course Planners, Kevin Tracey, MD, Cynthia Hahn, Emmelyn Kim, MPH, Tina Chuck, MPH have nothing to disclose.
- Phyllis W. Speiser, MD is the speaker and has nothing to disclose.
The need to be evidence-based

1. Wide variations in practice
   1. Continued use of ineffective treatments
   2. Excess use of inappropriate treatments
   3. Poor uptake of effective practice

2. Increasing consumerism
   1. Unvetted Internet information
   2. Direct-to-consumer pharma advertising

3. Increasing demand on $ resources
   1. Need to demonstrate efficacy

4. Exponential growth in research
   1. Need to compare & evaluate many studies’ variable quality & conflicting results.
Information overload

- MEDLINE 2012
  - >5,000 journals surveyed
  - 22,000,000 citations
Hierarchy of evidence for treatment decisions

- Meta-analysis of RCTs
- Systematic review of RCTs
- Individual RCT
- Observational studies
  - patient-important outcomes
- Basic research
  - test tube, animal, human physiology
- Clinical experience
Types of EBM studies

- **Diagnosis**
  - Selecting appropriate diagnostic tests
- **Therapy**
  - Selecting most effective/safest treatments
  - Cost-benefit
- **Prognosis**
  - Outcomes & complications
- **Associations/ Causes**
  - Identify etiologies: e.g., infectious, environmental, iatrogenic
What makes a review “systematic”?

• Based on a clearly formulated question
• Identifies relevant studies with pre-set criteria
• Appraises quality of studies
• Summarizes evidence by use of explicit methodology
• Recommendations are based on evidence gathered
Assessing quality in the medical literature

- Study design
- Type of intervention
- Endpoints examined
- Data analysis
Quality / validity of studies: Design

- **Study design**
  - Prospective v. retrospective
  - Cross-sectional v. longitudinal
  - Clinic population only v. case-control

- **Patient selection**
  - Consecutive v. nonconsecutive v. random
  - Age, racial, ethnic & gender balance
  - Power analysis to determine subject number?
  - Number of drop-outs
  - “Intention to treat”
Quality / validity of studies: Intervention

– Nature of intervention
  • Placebo-controlled v. best current treatment v. uncontrolled
  • Randomized or not
  • Blinded or not
  • Dose-ranging v. single dose

– Verification of methods
  • Same or different assays, inter- & intra- assay variability
  • Same or different endpoints
  • Empiric or historical normal reference data
  • Appropriateness of controls
Quality / validity of studies: Data

– Data collection
  • Prospective or retrospective
  • Intention to treat
  • Exclusion criteria for outliers
  • Compliance assurance (eg, weekly phone calls, patient diaries, pill counts, etc)

– Statistical analysis
  • Appropriateness of statistical methods
What is a meta-analysis?

Optional part of a systematic review

- Systematic reviews
- Meta-analyses
Meta-analysis: Are the studies consistent?

• Are variations in results between studies consistent with chance?

• If NO, then WHY?
  – Variation in population
  – Variation in study methods (biases)
  – Variation in intervention
  – Variation in outcome measure (e.g., timing)
Pitfalls of meta-analysis

- Potential bias in inclusion / exclusion criteria for study selection
  - Publication bias toward positive results
  - Keyword search
- Size
  - Number of studies
  - Sample size, total & individual
  - Attrition
- Length of follow-up
Pitfalls of meta-analysis, cont

• Methods of meta-analysis
  – Sensitivity analysis for robustness
    • Fixed vs random effects
    • Outlier exclusions
  – Stratification of subject populations

• Conclusions of meta-analysis
  – Weak if studies on opposite sides of Forest plot: “Heterogeneity”
GRADE system: Knowledge translation

- Transparent process of moving from evidence to recommendations
- Developed by representative group of international guideline developers
- Separates quality of evidence & strength of recommendations
- Stresses importance of outcomes of alternative management strategies
- Explicit acknowledgment of patients’ & providers’ values and preferences
- Clear, pragmatic interpretation of strong versus weak recommendations for clinicians, patients, and policy makers
Grading evidence

- **High quality**— Further research is very unlikely to change our confidence in the estimate of effect
- **Moderate quality**— Further research is likely to have an important impact on our confidence in the estimate of effect and may change the estimate
- **Low quality**— Further research is very likely to have an important impact on our confidence in the estimate of effect and is likely to change the estimate
- **Very low quality**— Any estimate of effect is very uncertain
## Strength of Recommendations

<table>
<thead>
<tr>
<th>Factor</th>
<th>High Rank</th>
<th>Low Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quality evidence</td>
<td>RCT</td>
<td>Case series</td>
</tr>
<tr>
<td>Balance of risk &amp; benefit</td>
<td>Low toxicity &amp; High efficacy</td>
<td>High toxicity &amp; High efficacy</td>
</tr>
<tr>
<td>Values &amp; preferences</td>
<td>Life-saving or QOL-enhancing</td>
<td>No major advance</td>
</tr>
<tr>
<td>Cost</td>
<td>Inexpensive</td>
<td>Costly</td>
</tr>
</tbody>
</table>
Evidence-based clinical decisions: Are antibiotics indicated in pediatric otitis media?

• Typical case: A 3 year old child with
  – Ear pain, low grade fever, irritability
  – Examination shows bilateral otitis media

• Should antibiotics be prescribed?
  – Benefits?
  – Risks?
Glasziou, Cochrane systematic review, 2003

- Systematic review of RCTs

- Question
  - Patients: children with otitis media
  - Intervention: antibiotics
  - Outcome: resolution of symptoms—WHEN?
  - Calculate odds ratios & confidence interval for each study & combine comparable data

- Comprehensive search

- Only 8 high quality studies (N= 2,287 children). These studies had:
  - Concealed randomization
  - Double blinding of treatments
  - Complete follow-up
Odds ratios allow comparisons of different studies in meta-analysis

<table>
<thead>
<tr>
<th></th>
<th>Test/Tx +</th>
<th>Test/Tx -</th>
</tr>
</thead>
<tbody>
<tr>
<td>Disease</td>
<td>A (TRUE POS)</td>
<td>B (FALSE NEG)</td>
</tr>
<tr>
<td>Unaffected</td>
<td>C (FALSE POS)</td>
<td>D (TRUE NEG)</td>
</tr>
</tbody>
</table>

Odds ratio for treatment efficacy =

AD / BC,

or TP x TN / FN x FP
Confidence intervals: Definition

Confidence intervals are based on the assumption that a study provides one sample of observations out of many possible samples that would be derived if the study were repeated many times.

For a 95% confidence interval, if the experiment were repeated many times, 95% of the intervals would contain the true treatment effect.
Endpoint #2: Pain at 2 – 7 days improved w/tx
Timing is important!

<table>
<thead>
<tr>
<th>Study</th>
<th>Odds Ratio (95% CI)</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Appelman</td>
<td>0.86 (0.34, 2.22)</td>
<td>121</td>
</tr>
<tr>
<td>Burke</td>
<td>0.65 (0.34, 1.22)</td>
<td>225</td>
</tr>
<tr>
<td>Damoiseaux</td>
<td>0.55 (0.32, 0.94)</td>
<td>240</td>
</tr>
<tr>
<td>Halsted</td>
<td>1.08 (0.39, 2.97)</td>
<td>89</td>
</tr>
<tr>
<td>Kaleida</td>
<td>0.50 (0.29, 0.85)</td>
<td>980</td>
</tr>
<tr>
<td>Mygind</td>
<td>0.45 (0.22, 0.90)</td>
<td>149</td>
</tr>
<tr>
<td>Thalin</td>
<td>0.57 (0.29, 1.10)</td>
<td>316</td>
</tr>
<tr>
<td>van Buchem (a)</td>
<td>0.43 (0.14, 1.27)</td>
<td>84</td>
</tr>
<tr>
<td>van Buchem (b)</td>
<td>0.57 (0.21, 1.56)</td>
<td>83</td>
</tr>
<tr>
<td>Pooled Estimate</td>
<td>0.57 (0.45, 0.73)</td>
<td>2287</td>
</tr>
</tbody>
</table>
## Effect of population size: Confidence interval and significance improve with larger or pooled samples

<table>
<thead>
<tr>
<th>Number in treatment arm</th>
<th>5</th>
<th>10</th>
<th>15</th>
<th>20</th>
<th>100</th>
</tr>
</thead>
<tbody>
<tr>
<td>Responders in treatment arm</td>
<td>4</td>
<td>8</td>
<td>12</td>
<td>16</td>
<td>80</td>
</tr>
<tr>
<td>Proportion responding in treatment arm</td>
<td>0.8</td>
<td>0.8</td>
<td>0.8</td>
<td>0.8</td>
<td>0.8</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Number in control arm</th>
<th>5</th>
<th>10</th>
<th>15</th>
<th>20</th>
<th>100</th>
</tr>
</thead>
<tbody>
<tr>
<td>Responders in control arm</td>
<td>2</td>
<td>4</td>
<td>6</td>
<td>8</td>
<td>40</td>
</tr>
<tr>
<td>Proportion responding in control arm</td>
<td>0.4</td>
<td>0.4</td>
<td>0.4</td>
<td>0.4</td>
<td>0.4</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>95% CI</th>
<th>-0.23 to 0.79</th>
<th>-0.04 to 0.71</th>
<th>0.04 to 0.67</th>
<th>0.10 to 0.64</th>
<th>0.27 to 0.52</th>
</tr>
</thead>
</table>

| p-value | 0.29 | 0.09 | 0.03 | 0.01 | <0.0001 |
Moral:

Any observed difference between two groups, no matter how small, can be made “statistically significant” - at any level of significance - by taking a sufficiently large sample.
Balanced decisions

- Are the differences are both clinically & statistically significant?
- Is the benefit greater than the risk?
- Failure to resolve pain at 2 to 7 days
  - 20% control
  - 13% antibiotics
- Absolute difference 7% in beneficial effect of antibiotics
  - Number Needed to Treat Effectively = 15
- Rates of antibiotics’ side effects
  - Increase in vomiting, rash, diarrhea 5%
  - Number Needed to Harm = 20
Subclinical hypothyroidism and ischemic heart disease: Winnowing publications for meta-analysis

Studies searched on PubMed, EMBASE and Cochrane Library using Medical Subject Heading database (n=2215)

Excluded (n=2133)
- Case-reports and letters (n=1123)
- Review articles (n=47)
- Not community-based populations (n=937)
- Intervention trials (n=26)

Full length articles reviewed (n=82)*

Further excluded (n=67)
- No report of thyroid status (n=8)
- No report of outcomes (n=22)
- Multiple publications (n=10)
- Not population-based (n=11)
- Editorials/Letters/Reviews (n=16)

Cross-sectional studies of SCH and IHD (n=6)

Longitudinal studies of SCH and IHD or cardiovascular mortality (n=9)
Forest plot of IHD prevalence in SCH and euthyroid controls: Age matters!

<table>
<thead>
<tr>
<th>Study</th>
<th>IHD prevalence (random)</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>01 Group 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hak (6)</td>
<td>2.30 [1.28, 4.13]</td>
<td></td>
</tr>
<tr>
<td>Inaizumi (13)</td>
<td>2.50 [1.13, 5.53]</td>
<td></td>
</tr>
<tr>
<td>Kvetny (18)</td>
<td>1.03 [0.73, 1.45]</td>
<td></td>
</tr>
<tr>
<td>Takashima (16)</td>
<td>1.46 [0.77, 2.75]</td>
<td></td>
</tr>
<tr>
<td>Tunbridge (34)</td>
<td>1.40 [0.96, 2.03]</td>
<td></td>
</tr>
<tr>
<td>Volzke (27)</td>
<td>1.27 [0.30, 5.39]</td>
<td></td>
</tr>
<tr>
<td>Walsh (14)</td>
<td>2.20 [1.20, 4.01]</td>
<td></td>
</tr>
<tr>
<td>Subtotal (95% CI)</td>
<td>1.57 [1.19, 2.06]</td>
<td></td>
</tr>
</tbody>
</table>

Test for heterogeneity: Chi² = 9.86, df = 6 (P = 0.13), I² = 39.1%
Test for overall effect: Z = 3.20 (P = 0.001)

<table>
<thead>
<tr>
<th>Study</th>
<th>IHD prevalence (random)</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>02 Group 2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cappola (8)</td>
<td>1.08 [0.85, 1.37]</td>
<td></td>
</tr>
<tr>
<td>Lindeman (11)</td>
<td>0.99 [0.63, 1.56]</td>
<td></td>
</tr>
<tr>
<td>Rodondi (9)</td>
<td>1.01 [0.79, 1.29]</td>
<td></td>
</tr>
<tr>
<td>van den Beld (10)</td>
<td>0.41 [0.05, 3.46]</td>
<td></td>
</tr>
<tr>
<td>Wilson (12)</td>
<td>0.82 [0.50, 1.34]</td>
<td></td>
</tr>
<tr>
<td>Subtotal (95% CI)</td>
<td>1.01 [0.87, 1.18]</td>
<td></td>
</tr>
</tbody>
</table>

Test for heterogeneity: Chi² = 1.68, df = 4 (P = 0.79), I² = 0%
Test for overall effect: Z = 0.13 (P = 0.90)

Total (95% CI)
Test for heterogeneity: Chi² = 19.97, df = 11 (P = 0.005), I² = 44.9%
Test for overall effect: Z = 2.20 (P = 0.03)
Evidence-based clinical decisions: Subclinical hypothyroidism

• “Only well-powered prospective randomized studies with age-stratified groups, and vascular events as the primary endpoint rather than surrogate markers, will give clear answers to this complex question” of whether & when to treat subclinical hypothyroidism.
EBM:
Beta blockers post MI-# needed to achieve desired outcome

- Prospective studies suggest that Mr. Jones' risk of death in the first year after his infarct is 8%
- A meta-analysis of RCTs of beta-blockers after MI suggests a 25% risk reduction
- Must treat 50 such pts to prolong a life
- Given the relatively small expense & low toxicity of generic beta-blockers, a trial of beta-blockers for Mr. Jones is clearly warranted
Thrombolytic therapy in MI: Power in N!
Importance of current data!

<table>
<thead>
<tr>
<th>Year</th>
<th>RCTs</th>
<th>Pts</th>
</tr>
</thead>
<tbody>
<tr>
<td>1960</td>
<td>1</td>
<td>23</td>
</tr>
<tr>
<td>1965</td>
<td>2</td>
<td>65</td>
</tr>
<tr>
<td>1970</td>
<td>4</td>
<td>316</td>
</tr>
<tr>
<td>1980</td>
<td>23</td>
<td>5767</td>
</tr>
<tr>
<td>1985</td>
<td>30</td>
<td>6346</td>
</tr>
<tr>
<td>1990</td>
<td>70</td>
<td>48154</td>
</tr>
</tbody>
</table>

Odds Ratio (Log Scale)

Favors Treatment  Favors Control

- P<.01
- P<.001
- P<.0001

Textbook/Review Recommendations

<table>
<thead>
<tr>
<th>Routine</th>
<th>Specific</th>
<th>Rare/Never</th>
<th>Experimental</th>
<th>Not Mentioned</th>
</tr>
</thead>
<tbody>
<tr>
<td>M</td>
<td>1</td>
<td>8</td>
<td>4</td>
<td>21</td>
</tr>
<tr>
<td>M</td>
<td>1</td>
<td>7</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>M</td>
<td>5</td>
<td>2</td>
<td>2</td>
<td>10</td>
</tr>
<tr>
<td>M</td>
<td>15</td>
<td>8</td>
<td>1</td>
<td>7</td>
</tr>
<tr>
<td>M</td>
<td>6</td>
<td>1</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Reviews and protocols for reviews on
The Cochrane Database of Systematic Reviews

2012: >5000 Cochrane reviews

Impact factor >5
Real world clinical decisions

Clinical state and circumstances

Research evidence

Patient values and preferences

Expertise
Knowledge Translation Research...

...study of the organization, retrieval, appraisal, refinement, dissemination, and uptake of knowledge (eg, important new knowledge from health research)
Knowledge Translation Research

Based on Hulley et al. Designing Clinical Research, 2007, p 23
Step 1. Generating Research Evidence

**Barrier**
- too little research addressing “real world” problems

**Solutions**
- large, simple randomized trials
- “head to head” comparisons
Examples

- Computerized decision support
- Evidence-based textbooks
- Evidence-based journal abstracts
- >57,000 EBM articles, 2009
- Systematic reviews
- Original journal articles
By the year 2020, 90% of clinical decisions will be supported by accurate, timely, and up-to-date clinical information and will reflect the best available evidence.

IOM Roundtable on Evidence-Based Medicine

This can’t happen without a better understanding of the barriers to translating knowledge into practice and ways to overcome them.

This can happen if $EBM+KT=0.90$
Clinical problem

• Previously healthy patient with flu wants to know:
  – Should he take Tamiflu?
  – Will it make him better quickly?
  – Is it worth the cost?
What’s the question?

• Is Tamiflu more effective than fluids, rest, and anti-pyretics?
• Does Tamiflu reduce flu severity?
• How much does Tamiflu cost?
How to find answers

- Search: Tamiflu AND efficacy
- Search: influenza AND Tamiflu AND Adults [limit to RCT in core clinical journals]
- Search: Tamiflu AND severity of illness
# Results

<table>
<thead>
<tr>
<th></th>
<th>SAE +</th>
<th>SAE-</th>
<th>Totals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rx</td>
<td>17</td>
<td>228</td>
<td>245</td>
</tr>
<tr>
<td>Pbo</td>
<td>19</td>
<td>110</td>
<td>129</td>
</tr>
</tbody>
</table>

Absolute risk reduction (ARR): $14.7\% \ (19/129 \ pbo) - 6.9\% \ (17/245 \ rx) = 7.8\%$

Relative risk reduction: $7.8\% \ (ARR)/14.7\% \ (pbo \ SAEs) = 53\%$

Number Needed to Treat: $1/7.8\% \ (ARR) = 13$ patients with flu need to be treated with tamiflu for 5 days to prevent one complication.

*JAMA 2000;283(8):1016-1024*
Informed decision

• For the physician: Integrate MD’s expertise with best published data

• For the patient: MD discusses:
  – Benefits
  – Risks
  – Cost
  – Individual values & preferences
Other resources:

1. Victor Montori, MD
   Mayo Clinic KER unit

2. McMaster Inst., Institute of Medicine

3. Online tutorial
   http://www.hsl.unc.edu/Services/Tutorials/EBM/welcome.htm

Summary: Evidence-based medicine (EBM)?

• Without EBM we are helpless in the face of
  – misguided experts
  – overenthusiastic experts
  – failure to report negative studies/ adverse outcomes
  – drug company hype

• Without EBM our ability is limited
  – to understand difficult tradeoffs
  – to help our patients make difficult decisions