**Evaluating Evidence**

**Practical Advice**

John Watkins, RPh, MPH

---

**EBM: A Practical Definition**

- When there is evidence of benefit and value, do it.
- When there is evidence of no benefit, harm or poor value, don’t do it.
- When there is insufficient evidence to know for sure, be conservative.


EBM is **NOT** Refusing to cover any drug or procedure that is not rigorously supported by Grade 1 evidence.

---

**Applying Population Based Medicine**

- Clinical practice guidelines, care pathways, algorithms
- Always visualize a hypothetical patient when considering a group
- Guidelines are not a substitute for applying good clinical judgment to the patient in front of you

---

**Why Evidence Based Medicine Was Needed**

Life before evidence-based medicine...

---

**Decision Making With EBM**

Clinician’s Expertise and Experience

Patient Treatment Preferences

Clinical Evidence

---

**Table**: Decision Making With EBM

| Decision Making | Clinical Evidence | Patient Treatment Preferences | Clinician’s Expertise and Experience | www.crbestbuydrugs.org | 6 |
We Have To Apply EBM in the Real World

Evidence-based medicine offers powerful tools but its usefulness has limits:
- Lack of evidence, lag time for outcome studies
- Who frames the questions
- Rigor vs. speed of review
- "Best Available Evidence"

Application
- How to apply the results to coverage decisions in a patient-specific, compassionate way
- How to make it clear, understandable
- Burden of proof always lies with the new technology

Applying EBM: Ethical Trade-offs

Justice:
- Equal access
- Fair procedures
- Consistent decisions
- "Fair equality of opportunity"

Autonomy:
- Freedom of choice
- Preference
- Sociocultural
- Religion
- Participative Decision making

Beneficence/nonmaleficence:
- Avoid risk without clear benefit
- Clinical quality
- "Basic minimum standard" of care
- Guidelines, practice standards

Practice Variation: Back Surgery

- Sevenfold variation between counties in Washington (rural areas highest)
- Maine study compared back surgery rates with results (residual pain and disability)
  - Areas with lowest surgery rates had best outcomes
  - Areas with highest rates had worst outcomes
  - Results adjusted for baseline (preop) severity
- Authors’ conclusion: surgeons who performed fewer procedures were better at identifying the best surgical candidates.
- More care does not necessarily improve quality or outcomes, but may improve patient satisfaction


Balancing Individual & Population Needs

- Fidelity—clinician’s duty to help the individual patient, and to be faithful to that one-on-one relationship
- Stewardship—clinician’s duty to help conserve resources in order to seek fairness to the population as a whole
- Clinicians must accept both responsibilities and balance them in daily practice—if they resist they can prevent the allocation process from working
- We need a common language to communicate the need for priority setting to patients and the public simply and honestly. If we do, they will accept the need for priorities.

Sabin JE. BMJ. 1998;347:1002

Putting a Human Face on Medical Necessity

Titus is a 2 year old recently diagnosed with globoid-cell leukodystrophy, a rare genetic disease with progressive CNS deterioration. The case fatality rate is 100%. His younger brother also has it but is still asymptomatic.

A case series (n=5) appeared in NEJM in 1998. All showed dramatic improvement following allogeneic stem cell transplantation, studied for as long as 9 years. The authors cite 6 earlier deaths which they attribute to older transplantation protocols. They have reviewed Titus' case and accepted him for this procedure.

The family’s HMO denied coverage saying that the procedure is investigational. It costs $500,000. You are an outside peer reviewer reading the appeal. What is your decision?

Leslie Knope, KOMO TV News, Seattle, November 30, 2004

Guidelines run amok: the Wernecke case

- Ed Wernecke
  - "All I wanted was the best treatment for her that would cause the least amount of side effects... When they give us the information, they need to give us alternatives."

- Kate Wernecke (13)
  - "I don't need radiation treatment. And nobody asked me what I wanted. It's my body!"

- Michelle Wernecke
  - "Let's take the time out to consider all the options. We're scared, let us absorb this, let's think about this."
Kayla Burt: Courage and High Tech Medicine

Jan 2003: Kayla announces she can no longer play intercollegiate basketball after surviving a cardiac arrest.

Aug 2004: Kayla announces she’s back, with an implanted defibrillator.

Winter 2003: A sidelined Kayla cheers her teammates.

2004-5: Career Season-- Everett Herald Sports Woman of the Year

V Foundation Comeback Athlete of the Year - 2005

2006: V Foundation Award Ceremony

Comparing Alternatives

**Defibrillator**
- $30-40,000 cost
- Lasts 7-10 years
- No outcomes in this patient population. May extend survival (but not certain)
- Improved functionality and QoL, possibly life saving
- Risks – normal surgical risks with the implantation procedure

**Iressa (Lung Ca)**
- $20,000
- 9 months
- No survival benefit in a large RCT
- Improved functionality and QoL over a few months at best
- Majority of pts elderly; even if remission achieved, # of years of life gained is not very great
- Subgroup is much more responsive

**Tarceva (Pancreatic)**
- $12,000
- 4 months (median)
- Median survival by 12.8 days
- Median incremental response rate by 0.7% vs gemcitabine alone
- Most patients middle-aged or elderly
- Not known whether any subgroup is more responsive

The Rest of the Story…

“I was just amazing for me to put on my uniform every night...I have no regrets about what I’ve done. I’m so grateful that I had the opportunity to play one more time. I’m at peace with the fact that it has to end.”

Genomic Success Stories: Tiffany House

- 23 y.o. female presented @ age 12, long H/O progressive skeletal muscle weakness, chronic respiratory infections, scoliosis, growth cessation
- Diagnosis (1994):
  - Juvenile onset Acid Malate Deficiency (Pompe)
  - Prognosis: death before age 20
- Physically unable to attend school by age 13
- Treatment:
  - 1999 – 1st late onset pt to get Myozyme on investigational protocol
  - Continues tx with empiric adjustment of dosing ($350,000/year)
- Response/current status:
  - Functionally improved, still wheelchair bound but gaining strength
  - BA from UTSA, magna cum laude 2005, now MA in 2007

http://www.amda-pompe.org/patients.htm

When Double-Blind Placebo Trials Are Not Ethically Appropriate …

“Parachutes reduce the risk of injury after gravitational challenge, but their effectiveness has not been proved with randomized controlled trials.”


Parachutes reduce the risk of injury after gravitational challenge, but their effectiveness has not been proved with randomized controlled trials.”

“Parachutes reduce the risk of injury after gravitational challenge, but their effectiveness has not been proved with randomized controlled trials.”


“Parachutes reduce the risk of injury after gravitational challenge, but their effectiveness has not been proved with randomized controlled trials.”

Evidence-Based Review Process

- Develop and refine key questions
- Gather evidence
- Review evidence
- Target a population
- Prepare evidence synthesis
- Draft response and/or drug monograph
- Review by peers and content experts

Step 1: Develop Key Questions

- What are they (or we) asking?
- Is this the best formulation of the question, or can we focus/improve it?
  - Why do you want to know this?
  - Is there a specific patient or type of patient behind the question?
  - Tell me more about this
- What questions must be answered before we can answer the original question?

Step 2: Gather the Evidence

- Where should we look for answers?
  - Primary Sources: Medline & other databases/Original Articles
    Unpublished Data
    FDA Documents (www.fda.gov/cder)
  - Secondary Sources: Cochrane and other databases
    Review articles, Formulary, P&T, Drugs, FMS
- What search strategy should we use?
- Backups: external drug information centers, medical librarians, etc.

Step 3: Review and Evaluate

Overview of Study Designs

<table>
<thead>
<tr>
<th>Higher Grade</th>
<th>Lower Grade</th>
</tr>
</thead>
<tbody>
<tr>
<td>START TYPE</td>
<td></td>
</tr>
<tr>
<td>PRIMARY ACTION</td>
<td>Start w/</td>
</tr>
<tr>
<td>CONCLUDING ACTIONS</td>
<td></td>
</tr>
</tbody>
</table>

Current best evidence from clinical care research

- Research articles should be appraised in a step wise approach

   Method Section
   Valid
   Evaluate the results
   Invalid

(Adapted from Stuart, M. and Strite, S. The Delfini Group.)
Look for Bias

- Observational
  “If I hadn’t believed it, I wouldn’t have seen it.”
  - Attributed to Yogi Berra
- Selection
  - Enrolling patients in studies
  - Patients dropping out of studies

Example: Prevalence of *H pylori* Colonization in U.S. Population

Postulated Incidence of *H. pylori* in U.S. Population

Actual Incidence of *H. pylori* in U.S. Population

All Clinical Trials are not equal!

- Statistical vs clinical significance
- Efficacy (trials) vs Effectiveness (real world)

What’s Your Opinion??

Which of the following are **clinically** significant?

- □ Mean HAQ-DI ↑ was 0.33 on drug, 0.56 on placebo.
- □ Average length of flu symptoms ↓ from 6 days to 5.5.
- □ Average life expectancy in cancer ↑ by one month.
- □ Average life expectancy in cancer ↑ by one year.
- □ % Responders (relief of pain for 2+ wks/month) was 56% with drug and 47% with placebo.
- □ Endoscopic healing of esophageal erosions at 8 weeks was 93% on Drug A, but only 86% on Drug B.
All Clinical Trials are not equal!

- Statistical vs clinical significance
- Efficacy (trials) vs Effectiveness (real world)
- Outcomes vs intermediate markers
- Secondary endpoints and post-hoc analysis

Outcomes vs Intermediates

- Meaningful outcomes:
  - Incidence of MI, stroke, death
  - 5-Year survival in cancer
- Relatively strong intermediates:
  - HbA1c
  - LDL-c
- Less reliable:
  - Blood pressure
  - CD4 counts in HIV

Primary vs Secondary Endpoints

- Primary Endpoints
  - Major goal of study
  - Designed to answer this question, with adequate statistical power
- Secondary Endpoints
  - Other endpoints of interest, but not main reason for doing the study
  - May not be adequately powered - use caution when interpreting
- Pitfalls of Post-Hoc Analysis

Marginally-Powered Studies and Secondary Endpoints: ELITE

The original ELITE Trial:
- Losartan vs captopril in ACE-naive CHF patients, >65 with NYHA II-IV CHF
- N = 722
- Endpoints:
  - Primary: Renal dysfunction progression
  - Secondary: all-cause mortality, CHF hosp. Admission
- Lancet 1997;349:747

ELITE Secondary Results

- 46% reduction in all-cause mortality with losartan vs captopril (p=0.035)
  - This was statistically significant (barely), but was a totally unexpected result!
- Combined endpoints:
  - 32% reduction in combined death or CHF admission (p=0.075 - not statistically significant)
  - 26% reduction in all-cause hosp adm (p=0.014 - this was statistically significant, but is too broad a combination. What does it mean clinically?)

ELITE II

- Larger follow-up study designed to confirm the secondary outcome results of ELITE I.
- N = 3152
- Patients > 60 c NYHA II-IV CHF
- Primary Endpoints:
  - All-cause mortality
  - Sub-classified: sudden death, CHF progression, MI, Stroke, Other CV
- Lancet 2000;355:1582
ELITE II Primary Results

- All-cause mortality:
  - 13% Higher for losartan (p=0.16 - not stat. sign.)
- Mortality + hosp admission
  - 25% Higher for captoril (p=0.08 - not sign.)
- No endpoints were significantly different!!

All Clinical Trials are not equal!

- Statistical vs clinical significance
- Efficacy (trials) vs Effectiveness (real world)
- Outcomes vs intermediate markers
- Secondary endpoints and post-hoc analysis
- Intention to treat analysis
- Generalizability (inclusion/exclusion criteria)

Generalizability

- Do the study patients look like the population we are interested in?
- If not, how far can we extrapolate the results?

Complete Patient Follow-up

- What happened to the patients enrolled in the study?
  - Patient is healthy and does not need to follow-up
  - Patient is ill and too sick to follow-up
  - Patient quit due to side effects or lack of benefit
- How are they accounted for in the data analysis?
- Rule of thumb
  - When <10% of enrolled patients are lost to follow-up the results are less likely to be effected

Step 4:

Target a Population
Relative vs Absolute Risk

- The fine print says that the 33% reduction was in a sub-population of CHF patients.
- The reference article says that if not treated, 3% of these patients will die within a year.
- A subgroup of these with Grade III-IV CHF had 15% mortality in the next year.
- Cardiosalve is flat-priced and costs about $1,000 per patient-year.

Is this product a good value? For Whom?

What is the Relative Risk Reduction?

\[ \text{RRR} = \frac{\text{incidence in untreated pts} - \text{incidence in treated}}{\text{incidence in untreated patients}} = 33\% \]

What is the Absolute Risk Reduction?

\[ \text{ARR} = \frac{\text{incidence in untreated pts} - \text{incidence in treated}}{\text{incidence in untreated}} \]

- incidence in untreated = 3%
- incidence in treated = 2%

\[ \text{ARR} = 3\% - 2\% = 1\% \]

What is the Number Needed To Treat to prevent one MI Death?

\[ \text{NNT} = \frac{1}{\text{ARR}} = \frac{1}{0.01} = 100 \]

So we must treat 100 patients with Cardiosalve for one year to prevent one MI death in this population.

Cost to prevent one death = $1,000 x 100 = $100,000!

The Power of Selecting Targeted Patients

Cost to prevent one event = $1,000 x 100 = $100,000!

**BUT:**
- If we target Grade III-IV CHF patients, the absolute risk is 5x higher (15% vs 3%)!
- That means the NNT = 20.
- Cost to prevent one event = $1000 x 20 = $20,000.

**Lesson:**
If we target a drug to a more selected population with a greater absolute risk, the NNT decreases and cost-effectiveness increases! This is a useful strategy.
Case Study: Xolair (Omalizumab)

- Humanized monoclonal IgE ab
  - IgE overproduced in allergic asthma
  - Xolair binds IgE
  - Blocks allergic response
- Unique asthma treatment
- Subcutaneous, every 2-4 wks
- Cost $10,000 per patient-year

Clinical Trial Results

Clinically significant benefit was limited to a small sub-population of patients:

- Primary endpoint: serious asthma exacerbations
  - Relative reduction 40-50%
  - BUT: only in 10-15% of patients
- Secondary endpoints
  - Asthma Hospitalizations, ER Visits—not really significant
  - Reduction in steroid use (1/10th the cost of Xolair)
  - Quality of life—marginal clinical significance

Summary

- Ask the right questions
- Do a thorough, careful literature search
- Read critically, be skeptical
- If adequate grade 1 evidence is lacking, weigh other information appropriately in balance
- Apply the results to the original question, using common sense and clinical judgment
- Follow up to be sure you’ve provided all the necessary information and there are no further questions