EBM, Study Design and Numbers

David Frankfurter, MD
Professor
OB/GYN
The George Washington University
1978?
1978 Best Picture of 1978 *The Deer Hunter*
Universal Studios

The 50th Academy Awards were held at the Dorothy Chandler Pavilion in Los Angeles, California on April 3, 1978. ... *Annie Hall*, winning 4 out of 5 nominations, including Best Picture.
Objectives

- Understanding of paradigms guiding scientific inquiry
- Explain principles of EBM
- Describe key features of study design and data analysis
Journey toward ultimate truth...
Paradigm Eras

• Pre-positivist (300 B.C. – late 17th Century)
  - Aristotle
  - Passive Observation used to explain the world
Paradigm Eras

• Positivist (late 17th Century- present)
  – John Stuart Mill
  – Scientific method
  – Fixed reference frame
  – Experimental models generalized to real world
  – Great catastrophes
Paradigm Eras

- **Post-positivist (1950s-present)**
  - Sir Karl Popper
  - Human knowledge is based *conjectures*
  - “Conjectures are a function of language, the observer, the subjects as well as the interaction between them”
  - Findings are relative
Philosophical basis of EBM

- Hierarchy of evidence which vary re quality
  - N of 1
  - Systematic review
  - RCT
  - Observational
  - Personal experience

- Evidence alone is insufficient
  - Perceived risks
  - Inconvenience
  - Patient values
  - Patient preferences

The Scientific

The Humanistic
Requirements

• In depth background knowledge
• Diagnostic skills
• Active listening
• Searching skill
• Critical reasoning
• Communication skills
The clinical question

- **P**opulation
  - What is the study population?

- **I**ntervention
  - What is the intervention/exposure of interest?

- **C**ontrol
  - What is the control population?

- **O**utcome
  - What is the outcome of interest?
n of 1 Study

• Single subject followed over time.
• Multiple blinded treatment assignments carried out:
  – Punctuated by washout period.

**Advantages**
  – Results in customization of intervention

**Disadvantages**
  – Costly
  – Time consuming
  – Cannot be generalized
Randomized Clinical Control Trial (RCT)

- Advantages
  - Gold Standard
  - Experiment in which the subjects are randomly assigned/allocated to an intervention versus placebo
  - Followed over the course of the study and the effect of the intervention is determined.
  - Reduces bias
    - Groups are normalized by randomization.
  - Can determine causation

- Disadvantages
  - Costly
  - Time consuming
  - Limited generalizability
Cohort

- Prospective study that can prove associations, not causation
- Advantages
  - Safe.
  - Cheaper than RCT.
  - Can show temporal associations.
    - Prospective or retrospective.
  - Standardized measurements.
  - Can look at variables that are not amenable to randomization (Smoking)
- Disadvantages
  - Expensive
  - May require long follow up time
  - How to find controls? How to adjust for confounders?
  - Blinding and randomization is also a challenge.
Case-Control

- Identify patients with outcome and then the exposure/treatment of importance
  - Retrospective

- Advantages
  - Better for studying rare disorders.
  - Cheap
  - Quicker

- Disadvantages:
  - More open to bias and confounders.
  - Selection of appropriate controls may be difficult
  - Can not come up with absolute risk, only with odds
Validity

- Randomization/concealment
  - Blinding
  - Balance
    - Anthropomorphic
    - Demographic
    - disease severity
    - prognostic characteristics
  - Be leery of post randomization subgroup analysis

- Treatment effect
  - Precision

- Similarity to my patients

- Intention to treat with adequate follow up
  - Accounting for all subject
Treatment Effect

- **Relative risk reduction (RRR)**
  - RRR 25%
    - Control event rate of 100% and experimental event rate of 75%?
    - Control event rate of 1% and experimental event rate of 0.75%?

- **Absolute risk reduction (ARR)**
  - Control event rate (CER) - experimental event rate (EER)
    - 100% - 75% = 25%
    - 1% - 0.75% = 0.25%
Number Needed to Treat (NNT)

- The number of patients that must receive an intervention in order to prevent 1 adverse or produce 1 positive outcome

<table>
<thead>
<tr>
<th>ARR</th>
<th>NNT</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.25 (25%)</td>
<td>4 (1 ÷ 0.25)</td>
</tr>
<tr>
<td>0.0025 (0.25%)</td>
<td>400 (1 ÷ 0.0025)</td>
</tr>
</tbody>
</table>
Oi vey!

- **p value**
  - Representation of the likelihood that the observed effect was the result of a random event

- **Type I error (α error)**
  - Rejecting the null hypothesis when it is true
    - No association exists between intervention and outcome
  - Treatment effect observed was a random event or result of bias
  - False positive

- **Type II error (β error)**
  - Accepting the null hypothesis when it is in fact false
    - An association exists between intervention and outcome
  - Assuming no treatment effect when one exists
  - Usually due to a sampling problem i.e. under powered
    - Inadequate sample size
  - False negative
### Confidence intervals

<table>
<thead>
<tr>
<th>Control Event Rate</th>
<th>Experimental Event Rate</th>
<th>Relative Risk Reduction (%)</th>
<th>Absolute Risk Reduction (%)</th>
<th>95% Confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>2/4</td>
<td>1/4</td>
<td>50</td>
<td>25</td>
<td>-174 to 92</td>
</tr>
<tr>
<td>10/20</td>
<td>5/20</td>
<td>50</td>
<td>25</td>
<td>-14 to 80</td>
</tr>
<tr>
<td>20/40</td>
<td>10/40</td>
<td>50</td>
<td>25</td>
<td>9.5 to 73</td>
</tr>
<tr>
<td>50/100</td>
<td>25/100</td>
<td>50</td>
<td>25</td>
<td>26.8 to 66.4</td>
</tr>
<tr>
<td>500/1000</td>
<td>250/1000</td>
<td>50</td>
<td>25</td>
<td>43.5 to 55.9</td>
</tr>
</tbody>
</table>
**Oi vey Redux**

<table>
<thead>
<tr>
<th>Disease</th>
<th>Present</th>
<th>Absent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive</td>
<td>a</td>
<td>b</td>
</tr>
<tr>
<td>Negative</td>
<td>c</td>
<td>d</td>
</tr>
</tbody>
</table>

- **Sensitivity**
  - Test’s ability to pick up all with a disease
    - \( \frac{a}{a+c} \)
- **Specificity**
  - Ability of a test to exclude those that do not have the disease
    - \( \frac{d}{b+d} \)
- **PPV**
  - \( \frac{a}{a+b} \)
- **NPV**
  - \( \frac{d}{c+d} \)
Pitfalls of Sensitivity and Specificity

- Assume sensitivity and specificity of 99%
- Disease prevalence of 1.0%
- What are the positive and negative predictive values of test results?

Test for chlamydia

<table>
<thead>
<tr>
<th>Disease</th>
<th>Present</th>
<th>Absent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive</td>
<td>9.9</td>
<td>9.9</td>
</tr>
<tr>
<td>Negative</td>
<td>0.01</td>
<td>980.1</td>
</tr>
</tbody>
</table>

Population of 1000
10 people have dz

PPV = a/a+b = 9.9/9.9 + 9.9 = 9.9/19.8 = 0.50 or 50.0%

Sens: 99%
a/a+c + 0.99=a/10 + 9.9=a - 0.01=c - 0.01=c

0.01=c
• Recommends against routine screening mammography in women aged 40 to 49 years

• Recommends biennial screening mammography for women between the ages of 50 and 74 years

AIM: USPSTF. 2009
• Against routine screening in ♀ 40 – 49
  - 50-59 need for fewer invitations for screening
• But…
  - RRR and CI same for 39-49 and 50-59

AIM: USPSTF. 2009
• 40-49 vs. 50-59 have similar
  - FN rate
  - FP rate
  - Biopsy frequency
  - Number of bx’s necessary to detect CA

Table 2. Age-Specific Screening Results From the BCSC

<table>
<thead>
<tr>
<th>Screening Result</th>
<th>40–49 y</th>
<th>50–59 y</th>
<th>60–69 y</th>
<th>70–79 y</th>
</tr>
</thead>
<tbody>
<tr>
<td>False-negative mammography result</td>
<td>1.0</td>
<td>1.1</td>
<td>1.4</td>
<td>1.5</td>
</tr>
<tr>
<td>False-positive mammography result</td>
<td>97.8</td>
<td>86.6</td>
<td>79.0</td>
<td>68.8</td>
</tr>
<tr>
<td>Additional imaging</td>
<td>84.3</td>
<td>75.9</td>
<td>70.2</td>
<td>64.0</td>
</tr>
<tr>
<td>Biopsy</td>
<td>9.3</td>
<td>10.8</td>
<td>11.6</td>
<td>12.2</td>
</tr>
<tr>
<td>Screening-detected invasive cancer</td>
<td>1.8</td>
<td>3.4</td>
<td>5.0</td>
<td>6.5</td>
</tr>
<tr>
<td>Screening-detected DCIS</td>
<td>0.8</td>
<td>1.3</td>
<td>1.5</td>
<td>1.4</td>
</tr>
<tr>
<td>Patients undergoing biopsy to diagnose 1 case of invasive breast cancer§</td>
<td>5</td>
<td>3</td>
<td>2</td>
<td>2</td>
</tr>
</tbody>
</table>

AIM: USPSTF. 2009
• Mammography screening for breast cancer
  - Sensitivity 90%\textsuperscript{a,b}
  - Specificity 95%\textsuperscript{a,b}

\textsuperscript{a}AIM: USPSTF. 2009
\textsuperscript{b}Mushlin et al. 1998 American Journal of Preventive Medicine
**Sensitivity of Mammogram**

<table>
<thead>
<tr>
<th></th>
<th>&lt;50</th>
<th>50-64</th>
<th>&gt;65</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Prevalence</strong></td>
<td>7/10,000&lt;sup&gt;a&lt;/sup&gt;</td>
<td>3/1,000&lt;sup&gt;a&lt;/sup&gt;</td>
<td>4.2/1000&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td><strong>PPV</strong></td>
<td>2.0%&lt;sup&gt;b&lt;/sup&gt;</td>
<td>6.0%&lt;sup&gt;b&lt;/sup&gt;</td>
<td>9%&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
</tbody>
</table>


<sup>b</sup>Calculated using SEER and USPSTF data
Women’s attitudes

• “USPSTF recognizes the benefit of screening seems equivalent for women aged 40-49 and 50-59 years...The USPSTF emphasizes the adverse consequences for most women—who will not develop breast cancer”**
  - 40-49 year olds
    • Need to screen 1000 to pick up 1 additional case
    • Risk of X-Ray exposure on future risk
    • Psychological impact of mammography
  - 50-59 year olds
    • Need to screen approx 500 to pick up 1 additional case
    • Risk of X-Ray exposure on future risk
    • Psychological impact of mammography

• Women with prior FP mammogram welcome repeat screening**

*AIM: USPSTF. 2009
** Baines et al. 1990 Cancer
Avoid being mislead

- The clinical question
  - Is it plausible
- Were the methods correct?
  - RCT
    - Randomization scheme
    - blinding
  - Cohort
    - Adequate control of confounders
  - Case control
    - Appropriate ascertainment of cases and controls
- Adequate follow up
- Assessment of all potential outcomes including harmful ones
  - Intention to treat