Lecture 1: Introduction to Epidemiology and Biostatistics

What is epidemiology and evidence-based medicine and why do I need to know about it?

Mathew J. Reeves BVSc, PhD
Associate Professor, Epidemiology

Medicine and the information age

- Medicine has become an information intense discipline
- The volume of new medical information is staggering
  - Example: MEDLINE (NLM, NIH)
  - Started in 1965
  - Contains 400,000 biomedical journals
  - Contains 13 million articles
  - 400,000 new articles are added each year (that's > 1,000 a day)
- Access to medical information has increased dramatically
  - Everyone is exposed to the media
  - Almost everyone has access to the internet
- Interest in medical information has increased exponentially
  - The media focus on "today's medical research breakthroughs"
  - Increased awareness and demand for patients and payers
  - Increasing demand for physician accountability

CHM's Information Management Curriculum

- Developed to address two major educational needs:
  - Medical Informatics and Critical appraisal
    - The 21st century physician needs to be able to find, process, evaluate, and integrate new information into clinical practice on an ongoing basis
    - This is the EBM Revolution.......
  - Changes to medical education (residency training)
    - All residents now require residents to demonstrate core competencies in clinical research and critical appraisal
    - (Practice-based learning)
    - Includes evidence-based medicine, quality improvement, and informatics
Need another reason?... The U.S. Health System is "challenged" on many fronts....

- Compared to other industrialized countries, the U.S. health care system....
  - Is more expensive: >$2.0 trillion/year, 16% GDP
  - Has the highest inflation (> 8%/yr)
  - Does not produce the best outcomes
  - Is not rated highly by its citizens or doctors
  - Does not cover all of its citizens.... ~15% uninsured (~46 Million)
U.S. Health Care: A high cost but low quality industry.

- Congressional Budget Office, 2007:
  - "The long-term fiscal condition of the US has largely been misdiagnosed. Despite the attention paid to the demographic challenges, such as the coming retirement of the baby-boom generation, our country's financial health will in fact be determined primarily by the growth in per capita health care costs."
  - Orszag and Ellis, NEJM, 357:1793, 2007
Projected Federal Spending for Medicare and Medicaid under Various Assumptions about the Growth Difference Between Health Care Costs and Per Capita GDP

How med money is spent

The US spent around $2.2 trillion this year on medical care.
Where the money went in 2021:

- Hospital care: 21%
- Physician, clinical services: 21%
- Prescription drugs: 10%
- Dental, other professional: 10%
- Administration: 7%
- Investment: 7%
- Nursing home: 6%
- Government public health activities: 3%
- Other medical products: 3%
- Home health care: 2%

Where does all the money go?

<table>
<thead>
<tr>
<th>Category</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hospital care</td>
<td>21</td>
</tr>
<tr>
<td>Physician, clinical</td>
<td>21</td>
</tr>
<tr>
<td>services</td>
<td></td>
</tr>
<tr>
<td>Prescription drugs</td>
<td>10</td>
</tr>
<tr>
<td>Dental, other</td>
<td>10</td>
</tr>
<tr>
<td>professional</td>
<td></td>
</tr>
<tr>
<td>Administration</td>
<td>7</td>
</tr>
<tr>
<td>Investment</td>
<td>7</td>
</tr>
<tr>
<td>Nursing home</td>
<td>6</td>
</tr>
<tr>
<td>Government public</td>
<td>3</td>
</tr>
<tr>
<td>health activities</td>
<td></td>
</tr>
<tr>
<td>Other medical</td>
<td>3</td>
</tr>
<tr>
<td>products</td>
<td></td>
</tr>
<tr>
<td>Home health care</td>
<td>2</td>
</tr>
</tbody>
</table>
Increased medical costs are...

- Driven primarily by the use of new medical therapies and technologies
  - many of which are not proven to be better or more cost-effective than existing treatments.
- Use of medical services is encouraged by the
  - fee-for-service model (rewards providers for delivering more care e.g., procedures and tests), and
  - lack of incentives for consumers to lessen their demand for services.
- Evidence that higher spending promotes better health outcomes and/or higher quality care is slim to none.

Regional Variation in Medicare Spending Per Capita, 2003

Orcutt and Etts. NEJM. 357:1799, 2007

Poor Quality Care

Institute of Medicine (IOM) Committee on the Quality of Health Care in America

  - "The current health care system frequently fails to translate knowledge into practice and to apply new technology safely and appropriately"
- Established 6 major aims for improving health care.
  Health care should be:
  - Safe, effective, patient-centered, timely, efficient, and equitable.
The importance of EBM in health system reform

- The development of a rational and sustainable health care system will need to be based in large part on the principles of EBM – including:
  - Comparative effectiveness
  - Costs and cost-effectiveness
  - Quantifying risks, and benefits
  - Patient utilities and preferences
  - Shared-decision making
  - Performance measures
  - Continuous quality improvement (CQI)

- Physicians need to understand these principles

Goals of the Epidemiology and Biostatistics Courses

- Epi-546 (SS 1st Year)
  - To provide a grounding in the principles of clinical epidemiology, and biostatistics (vocabulary, concepts, definitions, applications) that are fundamental to EBM.
  - 10 lectures

- Epi-547 (FS 2nd Year)
  - Small group sessions designed to further develop the concepts, definitions and applications of EBM, and to apply them in the evaluation of clinical studies (critical appraisal).

EBM vocabulary for 21st Century Medicine......

- Efficacy
- OR
- Cost-benefit
- RR
- NNH
- 95% CI
- Time-to-event analysis
- P-value
- Intention-to-treat
- Sensitivity
- ARR
- Likelihood ratio
- Population attributable risk
- RRR
- Effectiveness
- DB-PC-RCT
- Meta-analysis
- NNT
- HR
I. What is Epidemiology?

- Epi means "over all"
- Demos means "people"
- Epi + Demos = "All of the people"

- *Defn: The study of the distribution and determinants of disease*
- *Defn: The science behind disease control, prevention and public health*
- Epidemiologists plan, conduct, analyze and interpret medical research.

II. What is Evidence-Based Medicine?

- *Evidence-based medicine (EBM) is the conscientious, explicit and judicious use of the current best evidence in making decisions about the care of individual patients (Sackett 1996).*

- EBM is just one component of epidemiology and public health but it is the one that is the most relevant to you as a "doctor in training"....

**Relationship between Clinical Medicine, Public Health, EBM and Epidemiology....**

- Medicine
- Pub Health
- EBM
- EPI
- MDs trained in EBM
- Health professionals with PhD, MS, or MPH
EBM - Important Concepts

- Synthesis of individual clinical expertise and external evidence from systematic research.
- Stresses expertise in information gathering, synthesis and incorporation.
- De-emphasizes memorization.
- Relies on evidence-based medicine and on-line resources (e.g., PubMed, ACP Journal Club, Cochrane Database of Systematic Reviews).

EBM is concerned with every day clinical issues and questions

<table>
<thead>
<tr>
<th>Issue</th>
<th>Question?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal/abnormal?</td>
<td>Is the patient sick? What concerns are associated with disease?</td>
</tr>
<tr>
<td>Diagnosis</td>
<td>How do we make a diagnosis? How accurate are diagnostic tests?</td>
</tr>
<tr>
<td>Risk factors</td>
<td>What factors are associated with disease risk?</td>
</tr>
<tr>
<td>Prognosis</td>
<td>What is the likely outcome? What factors are associated with poor outcome?</td>
</tr>
<tr>
<td>Treatment</td>
<td>How does early detection improve outcome? Can we prevent disease?</td>
</tr>
<tr>
<td>Prevention</td>
<td>What factors result in the disease? What is the underlying pathogenesis?</td>
</tr>
</tbody>
</table>

Important points about Epi-546/7...

- Epi-546/7 seeks excellence not mediocrity both in the quality of its teaching and in the mastery of the subject by its pupils.
- Epidemiology requires "critical thinking"
- Epidemiology and biostatistics are not easy subjects—the concepts take time and require multiple explanations, exercises and discussions before they are mastered... I know this from personal experience.
- This subject therefore has a longer learning curve than most pre-clinical subjects.
However.... Epi-546 has its issues...

- It's only a 1 credit course
- It comes at you very fast. The final exam is just 4 weeks away.
- It's not easy - especially if you are not quantitatively orientated or have 'forgotten' how to think critically...
- But there is lots of help...
  - Online practice questions on Angel
  - Practice exams on Angel
  - Office hours
  - Supplemental help from CHM (if needed)
- But beware of blowing this course off and cramming for the final.... The evidence we have suggests that this is a high-risk strategy.

The 2009 Epi-546 Course

- There is a course pack - read it!
- Read the course policies carefully!
- All materials are also under the "Lessons" folder on Angel.
- Within each folder you will generally find:
  - .pdf of the core PowerPoint lectures slides
  - .pdf of course notes
  - Practice questions
  - Pre-recorded Camtasia lectures (some but not all topics)
- All "live lectures" will also be recorded and placed in the Live Lecture folder (there is no live lecture for Lecture 2: Descriptive Statistics)
- Note that most of the epi course materials are also available as open source documents on the Ech center web-site:
- http://learn.chm.msu.edu/epi/
The 2009 Epic:546 Course

- Instructor: John Doe
- Course pack (glossary, course notes, publications, PowerPoint slides)
- Assignments (including the written analysis and at least 1-2 essay questions)
- Required text (Fielder and Fielder) is designed to supply the concepts discussed in the lectures and covered in the course pack. Read it.

- Mid-terms exam - Wednesday Jan 31st 8-9 am, 102 Coward
  - Covers lectures 1-4
  - 15 questions (30% of total)
- Final exam - Monday Feb 9th 8-10 am, A133 LS
  - Covers all lectures 1-10
  - 25 questions

- About 25% of the test questions are multiple choice with the remainder being calculation based, fill in the blank, and/or short answer format.
- Pass = 344 (70%)
- This will be the first year I will be exploring the use of f-Checkers.

An exercise in critical thinking....

- Question:

  What is the evidence that attending lectures is beneficial?

What is the evidence that attending lectures is beneficial?

- Null hypothesis:
  - There is no association between lecture attendance (the exposure) and passing Epic:546 (the outcome)
- Exposure
  - Self-reported answer to the question “How many lectures did you go to?”
  - Dichotomized answers into:
    - Never (e.g., “None”, “none”)
    - Most (e.g., “all of them”, “most of them”)
- Outcome:
  - Final % score
  - Dichotomized into:
    - Fail (< 75%)
    - Pass (>= 75%)


Does it help to attend lectures to pass the Epi-546 course?

- **Data Collection:**
  - 20 subjects fail the final, 10 (50%) of whom attend a review session where they are asked:
    - "How many lectures did you go to?"
- **Results:**
  - 6/10 (60%) classified into the None group
- So, what should we do now?.....

---

Idea?...

---

All questions can be framed in terms of a 2 x 2 table (Describes relationship between Exposure and Outcome)

```
<table>
<thead>
<tr>
<th></th>
<th>Pass (95%)</th>
<th>Fail (&lt; 75%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ALL</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Exposure (Lectures)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>None</td>
<td>a</td>
<td>b</td>
</tr>
</tbody>
</table>
```

---
Case-control Study approach
- interview 10 students who passed exam (cases) and 10 who failed (controls)

<table>
<thead>
<tr>
<th>Lectures?</th>
<th>Outcome Pass (35%)</th>
<th>Outcome Fail (75%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALL</td>
<td>8</td>
<td>4</td>
</tr>
<tr>
<td>None</td>
<td>2</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>10</td>
</tr>
</tbody>
</table>

Calculate odds ratio (OR) = 8/2 = 6.0 (95% CI 0.81 - 44.3)

CCS Results
- Students who passed the exam were 6 times more likely to have attended lectures compared to those that failed.
- But potential problems of bias...
  - Selection bias amongst controls (only half the students who failed attended the review session)
  - Selection bias amongst cases (we don’t know if the 10 students who passed are representative of all the students who passed)
  - Recall bias (accuracy of reporting attendance)
  - Random error (small study)
    - 95% Confidence Interval (CI) for OR = 0.81 - 44.3

Alternative approach – a cohort study
- Collect data prospectively on attendance before the final exam
- But how to collect such data?... Ideas?....
- Imagine that 70% of the class were found to meet the definition of “ALL”.
- Now correlate this with the exam results....
Cohort Study approach
- follow all 100 students, 70% attend lectures

<table>
<thead>
<tr>
<th>Lecture?</th>
<th>All</th>
<th>Fall (&lt;70%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALL</td>
<td>65</td>
<td>5</td>
</tr>
<tr>
<td>None</td>
<td>15</td>
<td>15</td>
</tr>
</tbody>
</table>

\[ \text{relative risk (RR) = } \frac{0.92}{0.5} = 1.84 \]

Cohort Study Results

- Students who attended lecture were 1.84 times more likely to pass the exam than those that did not.

- Now, no potential problems of bias...
  - Selection bias is avoided because we studied everyone
  - Recall bias is avoided because we collected data prospectively
  - Study is larger and more precise (95% CI = 1.29 - 2.87)

- But imagine if we had gotten these results....

Alternative Cohort Results
- follow all 100 students, 70% attendance

<table>
<thead>
<tr>
<th>Lecture?</th>
<th>All</th>
<th>Fall (&lt;70%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALL</td>
<td>55</td>
<td>15</td>
</tr>
<tr>
<td>None</td>
<td>25</td>
<td>5</td>
</tr>
</tbody>
</table>

\[ \text{relative risk (RR) = } \frac{0.78}{0.83} = 0.94 \]

(95% CI = 0.77 - 1.15)
Alternative Cohort Study Results

- Students who attended lecture were 0.94 times less likely to pass the exam than those that did not.

- Why could this be a plausible finding? .......

Because of confounding .......

- Students who chose not to attend had greater baseline proficiency in epidemiology (due to prior education [MPH] or maybe higher IQ)

<table>
<thead>
<tr>
<th>Attendance</th>
<th>Exam success</th>
</tr>
</thead>
<tbody>
<tr>
<td>-</td>
<td>+</td>
</tr>
</tbody>
</table>

Baseline epi proficiency

What is the fix for this problem?