Why Take This Course?

- How to answer policy and management questions
  - What is our clinical pathways program doing? Should we keep it?
  - Should we fund the fear-based anti-smoking campaign?
- How to pose research question(s) which are relevant to a particular policy or management decision
  - How to ask research questions, not simply talk “about” something
- How to do your own research
- How to judge the “research” (including salesmen’s charts and casual) of others

Disciplines Differ in Research Methods

- Biostatistics emphasizes randomized trials (classical experiments)
- Epidemiology uses correlational analysis of observational data
  - Addresses causality through timing, “dose-response”, biological mechanism, …
- Social sciences emphasize the role of theory when analyzing observational data
  - Economics also uses “natural experiments”
  - Policy analysis also uses design of quasi-experiments

Why social sciences are different (1)

- Epidemiology developed for essentially biological questions
- In the social arena, people are making decisions themselves, often thinking ahead
- Example: “Christmas card sales (Granger) cause Christmas”
  - Christmas card sales generally precede Christmas
  - Granger causality is a formal statistical test for whether one event regularly precedes another
  - When humans decision-making is involved one event preceding another may not be good evidence for causality
Why social sciences are different (2)

- Randomized Clinical Trials
  - the gold standard for solving treatment selection bias problems
  - often not a practical, ethical or sometimes even possible alternative
- Consider a randomized trial to study the effect of divorce on children’s substance abuse
- Consider a randomized trial to see the effect of a new clinical pathways program in your hospital

HPM Application: Physician Profiling

- **Physician profiling**: statistical analysis of the treatment patterns of an individual physician or groups of physicians relative to other physicians
  - Management question: Does physician profiling result in lower costs or better health outcomes for our enrollees?
  - Policy question: Does physician profiling result in dumping of riskier patients?
  - Outcomes question: Does physician profiling improve health of population?
- Is an RCT possible? Necessary? Can you learn things in other ways?

Why do Disciplines Differ in Research Methods?

- Different methods are suited to different questions
  - Social sciences, policy and management offer differ
- These are difficult questions with no right answers
  - Even RCTs do not solve many problems
  - Can never really know if something is true
  - But you can sometimes know that something is not true
  - Very intelligent people often disagree
What’s new in this course?

- Using theory a lot
  - thinking through lots of possibly alternative theories
- Natural Experiments
- Designing Quasi-experiments
- Using panel data in new ways
- What is “bad” or misleading about randomized trials, even when they are possible
- We will also review ideas of experiments, but with a new goal:
  - Understanding elements of what makes an experiment useful in order to judge how well particular observational data, particularly natural experiments, can justify causal conclusions

Other Reasons to Take This Course

- Practical Implementation Skills
  - Data Management
  - Statistical Packages (SPSS)
  - Knowledge of Existing Datasets
  - Experience with Regression
- Risk Adjustment

Outline of Class

- Asking the Right Question
- Correlation and Causality
- An Example to Illustrate Both Points
  - Review of Cross-tabs and Chi-squared
- Experimental vs. Observational Data
- Generalizability

- The most important thing in this course:
  - Correlation vs. causality
  - Goal: You will all spot correlation/causality confusion in its many guises
Asking the Right Question

- Often interested in the effect of a particular action:
  - Raising cigarette taxes
  - Anti-smoking advertising campaign
  - Patient’s bill of rights legislation
  - New scheduling system in your clinic
  - New utilization management software
  - Paying physicians more for fewer specialist referrals
- What ideal research question addresses your policy question?

Correlation and Causality

- Correlation
  - In empirical data, when A is higher, B is systematically higher (or lower)
- Causality
  - A causes B
  - If we can somehow increase A, B will increase (or decrease)
- Correlation is not Causality
  - Something trained people know in principle but frequently mistake in practice

Example

- “Teenagers who do not get on well with their fathers are more likely to smoke, drink and use drugs than youngsters in average two-parent families”
- “A wake-up call for dads across America. Every father should look in the mirror and ask: ‘How often do I eat meals with my children? ‘Take them to religious services?’...’”
- “The study was based on a survey of teenagers”
What is the relevant policy question?

- What is the relevant policy question?

- What is the research question that was asked?

- Are research question and policy question the same? What links them?

Theory and The Policy Question

Theory:
1. Dad spends more time with teenager $\implies$ (2) Dad and teenager have better relationship $\implies$ (3) Teenager does not use drugs

If theory is true, then dads spending more time will result in less teenage drug use

Study addresses (2) and (3), not (1)

Counterfactuals

- Counterfactual: What would have happened if a different course of action had been taken
- Example: What would have happened to drug usage if Dad had spent more time with teenagers?
- Whenever someone advocates some course of action, there is an implicit counterfactual
Correlation: Cross Tabs Reviewed

<table>
<thead>
<tr>
<th></th>
<th>Good relationship with father</th>
<th>Poor relationship with father</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Uses Drugs</td>
<td>80</td>
<td>900</td>
<td>980</td>
</tr>
<tr>
<td>Does not use Drugs</td>
<td>720</td>
<td>300</td>
<td>1020</td>
</tr>
<tr>
<td>Total</td>
<td>800</td>
<td>1200</td>
<td>2000</td>
</tr>
</tbody>
</table>

Testing Correlation Reviewed

- To test for the statistical significance of the relationship… to see if the variables are actually associated
- Predict the number of counts that would occur if there was no association between drug usage and good relationships with Dad
- What would happen if drug use occurred at the same rate among those with good and bad relationships?
Expected Counts if no Association

<table>
<thead>
<tr>
<th>Good relationship with father</th>
<th>Poor relationship with father</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Uses Drugs</td>
<td></td>
<td>49%</td>
</tr>
<tr>
<td>49% x 800</td>
<td>49% x 1200</td>
<td>49%</td>
</tr>
<tr>
<td>392</td>
<td>588</td>
<td>49%</td>
</tr>
<tr>
<td>Does not use Drugs</td>
<td></td>
<td>51%</td>
</tr>
<tr>
<td>51% x 800</td>
<td>51% x 1200</td>
<td>51%</td>
</tr>
<tr>
<td>408</td>
<td>612</td>
<td>51%</td>
</tr>
<tr>
<td>Total</td>
<td>800</td>
<td>1200</td>
</tr>
</tbody>
</table>

Testing Correlation Reviewed:
Chi-square Test

\[ \chi^2 \text{ statistic } = \sum \frac{(\text{observed count} - \text{expected count})^2}{\text{expected count}} \]

\[ \chi^2 \text{ statistic } = \frac{(80-392)^2}{392} + \frac{(720-408)^2}{408} + \frac{(900-588)^2}{588} + \frac{(300-612)^2}{612} \]

\[ = 811 \]

Degrees of freedom = (#rows-1) x (#columns-1) = 1

VERY SIGNIFICANT!

Correlation: What does it tell us?

- Having a poor relationship with one’s father is CORRELATED with using drugs

Causality: Better relationships with fathers CAUSE less drug use
Theory and Causality

Theory:
(1) Dad spends more time with teenager $\implies$
(2) Dad and teenager have better relationship $\implies$
(3) Teenager does not use drugs

• We observe a correlation between (2) and (3)
• What else might cause the correlation?
  – Higher Dad education might explain better relationship and cause higher income which might in turn lower drug use.
  – One of TONS of explanations
  – What are some other possible explanations?

Other Explanations of Correlation
• A is correlated with B if:
  – A causes B
  – B causes A
  – C causes A and C causes B
• Suppose that
  – having a good marriage causes Dads to spend more time with their teenagers
  – having parents with good marriages causes teenagers to not use drugs

Would Dads with bad marriages spending more time with their teenagers lower drug usage?

Altered Example
• In a survey of teenagers it is found that teenagers who spend more time with their fathers are less likely to use drugs
• Therefore, advocates suggest fathers should spend more time with their teenagers in order to lower drug use

• Now what is the problem?
### Selection Bias

- Selection into the “treatment group” – spending more time with fathers – is not random
- Those factors that affect time spent with fathers (the “treatment”) may affect the likelihood of drug usage (the outcome)
- Selection Bias: the “treatment group” and “control group” differ in some way related to the outcome

### Theories of Selection Bias

- What are some theories for selection bias in this case?
  - Drug Use Causes Less Time with Dad
  - Other Factors Cause both Less Time and Drug Usage
- For social questions, it is essential to think through possible theories

### Randomized Experiments Solve Selection Bias

- Why?
  - We’ll review in more detail next time
- What experiment could you do here?
- What problems would there be with an experiment?
Generalizability

- Generalizability:
  - Do results from study apply to the real world?

- If randomized trial shows that increasing time with teenagers lowers drug use, does that mean it would be true in the country as a whole?

Take-home Lessons

- Clearly define the policy or management question that you would like to ask
- Compare the research question that you can ask with the question you would like to ask
- Think about all possible causal theories that explain observed correlations
- Randomized experiments solve the selection bias problem but can have generalizability problems
Cartoon: Questions for Homework

- Does this mean that a woman would be better off if she walked alone and in unfamiliar neighborhoods more often?
- What is the “policy question”?
- What is the implicit counterfactual?
- What is the relevant research question?
- How does the information given differ from the desired information?
- What experiment could answer the desired question?
- What are the generalizability issues associated with the experiment you describe?