Generalizing RCT results to broader populations

**IOM Workshop**
Washington, DC, April 25, 2013

Generalizing the right question, which is...?

Miguel A. Hernán
Harvard School of Public Health
Observational studies for CER

- Criticized because of potential for confounding and other biases
- Praised because of
  - faster results
  - lower cost
  - fewer ethical problems
  - better transportability (generalizability) of effect estimates to clinical practice than RCTs
Reasons for better transportability of observational estimates

- Patients similar to those in clinical settings
  - Patients in RCTs are highly selected
- Long follow-up allows to study clinical outcomes
  - Many RCTs rely on surrogate outcomes
- Realistic implementation of treatments/interventions
  - Often RCTs include intensive monitoring and additional measures to increase adherence
What if RCTs and observational studies had

- Same patients
- Same follow-up
- Same interventions

- Would they be answering the same question then?
  - Only if data analysis were the same
Consider RCT and observational study with similar design

- except that baseline treatment assignment is not randomly assigned in the observational study

- Example: an open-label randomized trial and an observational follow-up study to estimate the effect of postmenopausal hormone therapy on the risk of coronary heart disease

☐ The RCT is just a **follow-up study with baseline randomization**
Follow-up studies with and without baseline randomization

- Perhaps a more useful classification than RCTs and observational studies?
  - “every RCT is an OS on day 2”
  - Incidentally, I also wanted a T-shirt

- Think large simple trials and so-called “pragmatic” trials
  - Benefits of baseline randomization overshadowed by large noncompliance, loss to follow-up?
  - Typically not data collected to adjust for these biases
Follow-up studies with and without baseline randomization

- are often analyzed differently
  - Example: “intention to treat” analysis in randomized trial; “as treated” analysis in observational studies

- Why?
  - The analytic approach depends on the causal question (which is the same) and study design (which is almost the same)
Is it because of confounding adjustment?

☑ Observational studies need adjustment for baseline confounders
  ■ RCTs do not, at least when they are large

☑ But, other than adjustment for baseline confounding, **analysis should be identical**
  ■ both designs need adjustment for time-varying confounding and selection bias
  ■ because decisions after baseline are not randomly assigned under either design
The analytic approach is determined by the question

- We often compare the estimates of randomized and observational studies
  - which implies we believe they answer the same causal question
  - Example: the effects of statins and of postmenopausal hormone therapy on the risk of coronary heart disease

- What causal question is that?
  - Let us review a few possible questions
Causal questions in follow-up studies

1. The effect of being assigned to treatment, regardless of treatment received

- **Intention to treat (ITT) effect**
- Interventions to be compared:
  - be assigned to treatment A at baseline and remain in the study until it ends
  - be assigned to treatment B at baseline and remain in the study until it ends
- Requires adjustment for time-varying selection bias due to loss to follow-up
Causal questions in follow-up studies

2. The effect of receiving the treatment regimes specified in the study protocol

- **Per protocol** effect

- Example of interventions to be compared:
  - receive treatment A continuously between baseline and study end (unless toxicity arises)
  - receive treatment B continuously between baseline and study end (unless toxicity arises)

- Requires adjustment for time-varying confounding and selection bias
Causal questions in follow-up studies

3. The effect of receiving treatment regimes other than the ones specified in the study protocol

- Related to \textit{as treated} effects
- Example of interventions to be compared:
  - receive A as per protocol
  - receive B but switch from B to A if LDL-cholesterol raises above 160 mg/dL (4.1 mmol/L)
- Requires adjustment for time-varying confounding and selection bias
ITT, per protocol effects vs. ITT, per protocol analyses

- Typical ITT, per protocol, and as treated analyses of randomized trials
  - Usually do not adjust for time-varying (post-baseline) variables
  - Possibly biased estimates of ITT, per protocol, and as treated effects (Robins et al, since 1986)

- Warning: Instrumental variable analyses require separate discussion
Efficacy and effectiveness

- These terms are probably useful in simple settings with short-term interventions, but ambiguous in complex settings with sustained interventions over long periods.

- An explicit definition of the interventions that define the causal effect of interest is more informative.
  - ITT effect does not necessarily measure *effectiveness* in the real world.
  - Per protocol effect may measure *effectiveness* but does not generally measure *efficacy*
Observational studies analyzed like randomized trials...

- We need to define the observational analogs of:
  - intention to treat
  - per protocol or as treated effects to know what question to generalize

- Hint: “New user” designs are a first step towards estimating ITT effects in observational studies
... and vice versa

- Because the analysis of many RCTs require adjustment for time-varying variables to estimate:
  - intention to treat
  - per protocol
  - as treated

causal effects

- Hint: Adjustment for time-varying variables may require techniques like inverse probability weighting or the g-formula
  - Robins 1986
Conclusions

1. Question of interest must be clearly specified
   - ITT, per protocol, other?
   - In both RCTs and observational studies

2. Analysis of observational studies and RCTs must be the same
   - Except adjustment for baseline confounding

3. The terms *effectiveness* and *efficacy* are too vague to define causal questions
Supplementary materials
4 examples of follow-up studies with and without randomization

1. Classic cohort study
   - Nurses’ Health Study
   - Postmenopausal hormone therapy and coronary heart disease (CHD)

2. Randomized trial
   - Women’s Health Initiative
   - Postmenopausal hormone therapy and breast cancer

3. Electronic medical records
   - THIN
   - Statins and coronary heart disease (CHD)

4. Claims database
   - USRDS Medicare
   - Epoetin and mortality
Example #1

- Observational study analyzed like a randomized experiment

- Question: What is the effect of postmenopausal hormone therapy on risk of coronary heart disease in postmenopausal women?

- Data: Nurses’ Health Study
  - Hernán et al. *Epidemiology* 2008
Answers

☐ Observational studies
   ■ >30% **lower risk** in current users compared with never users
   ○ e.g., HR 0.68 in Nurses’ Health Study (Grodstein et al. *J Women’s Health* 2006)

☐ Randomized trial
   ■ >20% **higher risk** in initiators compared with noninitiators
   ○ HR 1.24 in Women’s Health Initiative (Manson et al. *NEJM* 2003)
**WHI: ITT effect estimates**

**Hazard ratio (95% CI) of CHD**

<table>
<thead>
<tr>
<th>Category</th>
<th>Hazard Ratio (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall</td>
<td>1.23 (0.99, 1.53)</td>
</tr>
<tr>
<td>Years of follow-up</td>
<td></td>
</tr>
<tr>
<td>0-2</td>
<td>1.51 (1.06, 2.14)</td>
</tr>
<tr>
<td>&gt;2-5</td>
<td>1.31 (0.93, 1.83)</td>
</tr>
<tr>
<td>&gt;5</td>
<td>0.67 (0.41, 1.09)</td>
</tr>
<tr>
<td>Years since menopause</td>
<td></td>
</tr>
<tr>
<td>&lt;10</td>
<td>0.89 (0.54, 1.44)</td>
</tr>
<tr>
<td>10-20</td>
<td>1.24 (0.86, 1.80)</td>
</tr>
<tr>
<td>&gt;20</td>
<td>1.65 (1.14, 2.40)</td>
</tr>
</tbody>
</table>
Randomized trial estimated the ITT effect

- What is the CHD risk in women assigned to **initiation** of hormone therapy compared with women assigned to no initiation of hormone therapy?

- Design and analysis:
  - Women randomly assigned to initiation of hormone therapy or placebo
  - Analytic approach: Compare risk between initiators (**incident** users) and nonusers of hormone therapy
Observational studies did not estimate ITT effect

- What is the CHD risk in women who are currently taking hormone therapy compared with women who are not?

- Design and analysis:
  - Women are asked about therapy use
  - Analytic approach: Compare risk between prevalent users and nonusers of hormone therapy (current vs. never)
“Current vs. never” contrast does not address any relevant question

☐ Consider a woman wondering whether to start hormone therapy
  ■ The current vs. never contrast does not provide the information she needs

☐ Consider a woman wondering whether to stop hormone therapy
  ■ The current vs. never contrast does not provide the information she needs

☐ Nothing interesting to generalize!
What if observational data used to estimate analog of ITT effect?

- Use observational data to answer same question as randomized trial
  - Comparing the risk in incident users vs. nonusers

- Re-analyze observational studies to estimate the observational analog of the ITT effect
## ITT effect estimates

<table>
<thead>
<tr>
<th></th>
<th>WHI</th>
<th>NHS</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Overall</strong></td>
<td>1.23 (0.99, 1.53)</td>
<td>1.05 (0.82, 1.34)</td>
</tr>
<tr>
<td><strong>Years of follow-up</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0-2</td>
<td>1.51 (1.06, 2.14)</td>
<td>1.43 (0.92, 2.23)</td>
</tr>
<tr>
<td>&gt;2</td>
<td>1.07 (0.81, 1.41)</td>
<td>0.91 (0.72, 1.16)</td>
</tr>
<tr>
<td><strong>Years since menopause</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;10</td>
<td>0.89 (0.54, 1.44)</td>
<td>0.88 (0.63, 1.21)</td>
</tr>
<tr>
<td>10-20</td>
<td>1.24 (0.86, 1.80)</td>
<td>1.13 (0.85, 1.49)</td>
</tr>
<tr>
<td>&gt;20</td>
<td>1.65 (1.14, 2.40)</td>
<td>--</td>
</tr>
</tbody>
</table>
When **same question** is asked

- No shocking observational-randomized discrepancies for ITT estimates
  - though wide CIs in both studies

- Any residual confounding?
  - Probably, but insufficient to explain the original discrepancy
But ITT analyses are problematic

Hernán, Hernández-Díaz. *Clinical Trials* 2012

- ITT effect affected by adherence
  - Imperfect adherence in both randomized and observational studies
- ITT inappropriate for safety outcomes
- We also estimated *per protocol* effect
  - Inverse probability (IP) weighted analyses to adjust for “noncompliance”
  - Again no randomized-observational discrepancies
- Toh et al. *Ann Intern Med* 2010
EXAMPLE #2

☐ Randomized experiment analyzed like an observational study

☐ Question: What is the effect of postmenopausal hormone therapy on risk of breast cancer in postmenopausal women?

☐ Data: Women’s Health Initiative randomized clinical trial
  ■ Toh et al. *Epidemiology* 2010; 21:528-539
Methodologic approach to estimate per protocol effect

- Estimate IP weights to adjust for time-varying confounding
  - Need data on post-randomization variables
- Estimate IP weighted hazards model to estimate
  - Hazard ratios
  - Survival (or cumulative incidence)
- Compare survival curves for continuous treatment vs. no treatment
  - Standardize curves to baseline variables
Hazard ratio of breast cancer
Hormone therapy vs. placebo

- Intention to treat analysis
  - 1.25 (1.01, 1.54)

- Per protocol analysis (IP weighted)
  - 1.68 (1.24 to 2.28)

- Suppose you are a woman considering initiation of hormone therapy and who plans to take it as instructed by your doctor
  - Which hazard ratio do you want?
EXAMPLE #3
Electronic medical records

☐ Question: What is the effect of statin therapy on CHD risk?

☐ Data: UK THIN (electronic medical records)
  ■ ~75,000 eligible patients
  ■ Used to emulate a sequence of observational “trials” of statin initiation
  □ Generalization of new-user design
  ■ Danaei et al. *Statistical Methods in Medical Research* 2013
The Changing Face of Epidemiology

With Great Data Comes Great Responsibility

Publishing Comparative Effectiveness Research in Epidemiology

Miguel A. Hernán

Comparative effectiveness research has been enshrined in the US Healthcare Reform Law of 2010. The law mandates the creation of a Patient-Centered Outcomes Research Institute (PCORI), which will establish national research priorities and methodological standards, and will carry out research. The UK’s National Institute for Health and Clinical Excellence, set up in 1999, was the world pioneer in this area. Though the organizational structure and duties of the American and British Institutes vary (eg, the US Institute is barred by law from considering the cost-effectiveness of interventions), both institutes have an overarching common goal: to improve the public’s health through research on the relative effectiveness of different interventions. These interventions include medical treatments, changes in health care organization and delivery, community and workplace interventions, individual lifestyle modifications, etc.

Upon first hearing the above, many epidemiologists quickly retort: “Isn’t comparative effectiveness research something we have always done under different names?” The answer is yes, of course. Epidemiologists are natural comparative effectiveness researchers. In fact, the new US law stipulates that the Board of Governors of the Institute shall collectively have scientific expertise in “epidemiology, decision sciences, health economics, and statistics.”
CONSORT
flowchart of emulated “trials”

32.9 million potential person-trials
(617,432 patients)

- 3.2 million prior statin use or <2 years on database
- 12.5 million major chronic diseases
- 16.4 million no recent data on confounders

844,800 eligible person-trials
(74,806 patients)

- 3.2 million prior statin use or <2 years on database
- 12.5 million major chronic diseases
- 16.4 million no recent data on confounders

13,599 initiators

- 245 died
- 62 lost to follow-up
- 117 cases
- 13,175 alive and event free at the end of follow-up

831,201 non-initiators

- 16,094 died
- 2,894 lost to follow-up
- 6,218 cases
- 805,995 alive and event-free at the end of follow-up

5/7/2013 Observational - Randomized
Hazard ratio (95% CI) of CHD THIN “trials” 2000-2006

<table>
<thead>
<tr>
<th></th>
<th>Intention-to-treat analysis</th>
<th>Per-protocol analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unique cases</td>
<td>635</td>
<td>488</td>
</tr>
<tr>
<td>Unique persons</td>
<td>74,806</td>
<td>74,806</td>
</tr>
<tr>
<td>Cases</td>
<td>6,335</td>
<td>4,849</td>
</tr>
<tr>
<td>Person-“trials”</td>
<td>844,800</td>
<td>844,800</td>
</tr>
<tr>
<td>Age-sex adjusted</td>
<td>1.29 (1.06, 1.56)</td>
<td>1.54 (1.09, 2.18)</td>
</tr>
<tr>
<td>Adjusted for covariates</td>
<td>0.89 (0.73, 1.09)</td>
<td>0.84 (0.54, 1.30)</td>
</tr>
<tr>
<td>Adjusted for covariates (excluding first year of follow-up)</td>
<td>0.71 (0.53, 0.94)</td>
<td>0.53 (0.27, 1.02)</td>
</tr>
</tbody>
</table>

5/7/2013 Observational - Randomized
What if we had compared prevalent (not incident) users vs. nonusers?

- **Current users**
  - HR: 1.42 (1.16, 1.73)

- **Persistent (1 yr) current users**
  - HR: 1.05

- **Persistent (2 yrs) current users**
  - HR: 0.77 (0.51, 1.18)

- We can get any result we want by changing the definition of current user!
  - Confounding-selection bias tradeoff

- **Nothing interesting to generalize here!**
  - See also Danaei et al. *Am J Epidemiol* 2012
Question: What is the effect of different doses of epoetin therapy on the mortality risk of patients undergoing hemodialysis?

Data: US Renal Data System (Medicare claims database)
- ~18,000 eligible elderly patients
- Zhang et al. CJASN 2009; 21:638-644
ITT effect not interesting here

- Because epoetin dose and use usually changes every month
- If we assigned individuals to whatever dose they are receiving at baseline
  - Adherence too low
  - Effect greatly attenuated
- Therefore “per protocol” analysis only
Survival under 3 epoetin dosing regimes

References

RCTs

References

Observational studies - Methods


- Hernán MA, Robins JM. Observational studies analyzed like randomized experiments: Best of both worlds. Epidemiology 2008; 19(6):789-792


References
Observational studies - Applications


