Analyzing observational studies like randomized trials, and vice versa

Beyond the new-users design

ICPE 2012, Barcelona, Spain

Miguel A. Hernán
Harvard School of Public Health
Did you notice that...

randomized trials and observational studies are analyzed differently?

☐ Do you wonder why?
Method of analysis determined by the clinical question

- If randomized and observational studies ask the same questions, they should use the same analysis.

- Randomized and observational studies should ask the same questions.
  - otherwise we could never compare their answers.
Exception: adjustment for baseline confounders

- Observational studies need adjustment for baseline confounders
- Randomized trials do not
  - At least when they are large
- But, other than that, analysis should be identical
  - Both observational and randomized studies need adjustment for time-varying confounders
If different analyses

At least one does not correspond to the question of interest

☐ Are we asking the correct questions in observational studies?

☐ Are we asking the correct questions in randomized trials?
Asking the right questions

A first step towards getting the right answers in epidemiologic research

Miguel A. Hernán
Harvard School of Public Health
A young couple moves into an apartment and decides to repaper the dining room. They ask the neighbor who has a dining room the same size,

“How many rolls of wallpaper did you buy when you papered your dining room?”

“How many rolls of wallpaper did you buy when you papered your dining room?”

“So”, he says.

So the couple buys seven rolls of expensive paper, and they start papering. When they get to the end of the fourth roll, the dining room is finished. Annoyed, they go back to the neighbor and say

“We followed your advice, but we ended up with three extra rolls!”

“So”, he says, “that happened to you too.”

(Cathcart and Klein. *Plato and a Platypus Walk into a Bar...*)
4 examples

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
Classic cohort study

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

- Overall: 1.23 (0.99, 1.53)
- Years of follow-up:
  - 0-2: 1.51 (1.06, 2.14)
  - >2-5: 1.31 (0.93, 1.83)
  - >5: 0.67 (0.41, 1.09)
- Years since menopause:
  - <10: 0.89 (0.54, 1.44)
  - 10-20: 1.24 (0.86, 1.80)
  - >20: 1.65 (1.14, 2.40)
Why did observational studies get it “wrong”?

- Popular theory: residual confounding
  - insufficient adjustment for lifestyle and socioeconomic indicators
  - Corollary: causal inference from observational data is a hopeless undertaking

- An alternative: Observational and randomized studies asked different questions
Randomized trial asked a clinically relevant question

- What is the CHD risk in women who initiate hormone therapy compared with women who do not?

- Design and analysis:
  - Women randomly assigned to initiation of hormone therapy or placebo
  - Analytic approach: Compare risk between incident users and nonusers of hormone therapy (Intention-to-treat)
Observational studies did not ask a relevant question

- 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 a 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
Our strategy

- 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
  - Hernán et al. Biometrics 2005
  - Hernán et al. Epidemiology 2008
Data: The Nurses’ Health Study

- Observational follow-up study
  - ~80,000 women with diet, lifestyle data in 1980
- Lifestyle and health information updated by questionnaire every two years
  - Use of hormone therapy
  - Diagnosis of CHD (confirmed by physician)
  - Risk factors for CHD
- We use this observational study to emulate a “trial” of hormone therapy
Protocol of the NHS “trial”
Interventions and Eligibility criteria

- Treatment regimes
  1) Initiation of oral estrogens plus progesterone therapy at baseline
  2) No hormone initiation at baseline

- Similar eligibility criteria as randomized experiment
  - Including washout interval: no hormone use in 2-year period before baseline
Protocol of the NHS “trial”
Baseline and Follow-up

☐ Baseline:
- Initiators: month of initiation in 2-yr period before the 1984 questionnaire
- Non initiators: average baseline month among initiators in same period

☐ Follow-up
- From baseline to CHD diagnosis, death from other causes, loss to follow-up, or June 2000, whichever came first
The NHS “trial”

Summary

- The NHS “trial” can be viewed as a nonrandomized, nonblinded trial that mimics the eligibility criteria, definition of start of follow-up, and treatment arms of the WHI randomized trial.

- Some differences
  - Distribution of baseline characteristics
  - Shorter time since menopause in NHS
  - Longer follow-up in NHS than in WHI
Protocol of the NHS “trial”

Intention to treat (ITT) principle

☐ Compare the CHD risk between initiators and noninitiators of hormone therapy at baseline

☐ Regardless of future use during the follow-up

☐ Observational analog of the ITT effect
  - Using Cox model like the WHI did
The NHS “trial”
Non randomized after all

- All confounders have to be appropriately measured and adjusted for in the model
- We included the following baseline variables
  - Age, past hormone use, parental history of myocardial infarction before 60y, education, husband’s education, ethnicity, age at menopause, calendar time, high cholesterol, high blood pressure, diabetes, angina, stroke, coronary revascularization, osteoporosis, body mass index, cigarette smoking, aspirin use, alcohol intake, physical activity, diet score, multivitamin use, and fruit/vegetable intake
Note:
No “fancy” methods necessary

- Just a Cox model
  - Not even time-varying variables

- But question has been changed
  - from comparison of current (prevalent) users vs. never users
  - to comparison of initiators (incident users) vs. noninitiators of hormone therapy
One more thing:
A sequence of NHS “trials”

- We started the NHS “trial” during the period before the 1984 questionnaire but there is nothing special about the 1984 questionnaire
- We also started a “trial” in the periods before the 1986, 1988, ... 1998 questionnaires
  - A sequence of nested “trials”
- Each woman may participate in a maximum of 8 trials
The NHS “trials”

☐ For each trial
  ■ Follow-up started at the trial-specific baseline and ended at diagnosis of CHD, death, lost to follow-up, or June 2000
  ■ Eligibility criteria applied at baseline

☐ We pool data across “trials” to obtain an effect estimate with a narrower confidence interval
  ■ Robust variance because of within-subject correlation
Results
Women eligible for NHS “trials”

☐ 34,472 women contributed to trials
  ■ 1,021 CHD cases

☐ Pooling over “trials”
  ■ On average, each woman participated in 4.4 trials
  ■ 152,479 participants
  ■ 6,602 initiators
  ■ 3,597 CHD cases
<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
  - though wide CIs in both studies

- What about the popular hypothesis?
  Any residual confounding?
  - Probably, but insufficient to explain the original discrepancy
Clarification:

- In the interest of time, I simplified Example #1 to make a simple point
  - *Question matters more than methods*

- Full analysis is more nuanced
  - “Fancy” methods used
  - Here I presented “intention to treat” estimates only for comparison with WHI
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 conducted IP weighted analyses to adjust for “noncompliance”
  - Per protocol analysis
  - No randomized-observational discrepancies
EXAMPLE #2
Randomized Trial

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

☐ Data: Women’s Health Initiative randomized clinical trial
  ■ ~16,000 postmenopausal U.S. women
  ■ Toh et al. *Epidemiology* 2010; 21:528-539
Effect of hormone therapy, what effect?

☐ Effect of assignment to hormone therapy under the study’s conditions?
  - Intention to treat effect

☐ Effect of hormone therapy use as instructed by the study’s protocol?
  - Per-protocol effect

☐ BOTH
  - They answer different questions
Methodological challenges for per-protocol effect

☐ Time-varying treatment
  ■ Women may not adhere to their assigned treatment (hormone therapy or placebo)

☐ Time-varying confounders
  ■ Use of hormone therapy depends on age, BMI, symptoms...
  ■ may be affected by prior treatment

☐ Also better to estimate absolute risks
  ■ Appropriately adjusted survival curves
  ■ Not only hazard ratios
Methodologic approach for per-protocol effect

- Estimate **per-protocol effect**
  - Continuous treatment vs. no treatment
- Inverse probability (IP) weighting to adjust for time-varying confounding
- Hazards model to estimate time-varying hazards
- Standardization of the corresponding survival curves
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 plan to take it as instructed by your doctor
  - Which hazard ratio do you want?
% free of breast cancer under full adherence to assigned treatment

Toh et al. *Epidemiology* 2010; 21:528-539 (w/ SAS programs)
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-users design
  ■ Danaei et al. *Statistical Methods in Medical Research* 2011
Statistical analysis

- Observational analogs of “intention to treat” and “per protocol” analyses
  - Appropriately adjusted for baseline and time-varying confounders via IP weighting

- Potential confounders
  - Sex, age, LDL-cholesterol, HDL-cholesterol, BMI, smoking, alcohol use, systolic blood pressure, diabetes, hypertension, atrial fibrillation, use of antihypertensives, insulin, other lipid-lowering drugs, and beta-blockers, doctors visits, referrals, hospitalizations in last 3 months, etc.
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)

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
Adherence to treatment

![Graph showing adherence to treatment over months of follow-up, comparing Non-initiators and Initiators.](image)
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</td>
<td>0.71 (0.53, 0.94)</td>
<td>0.53 (0.27, 1.02)</td>
</tr>
<tr>
<td>(excluding first year of follow-up)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

23-Aug-12 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

- Let’s see another piece of data
Mortality hazard ratio for statins in CHD secondary prevention studies

- RCTs: 0.84 (0.77, 0.91)
- Observational studies
  - Incident users: 0.77 (0.65, 0.91)
  - Prevalent-incident mix: 0.70 (0.64, 0.78)
  - Prevalent users: 0.54 (0.45, 0.66)

- Danaei et al. *Am J Epidemiol* 2012
EXAMPLE #4
Health claims database

- 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
“Intention to treat” 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
Methodological challenge

- Time-varying treatment
  - Use and dose of epoetin varies over the course of the disease

- Time-varying confounders
  - Hematocrit level, comorbidities
  - may be affected by prior treatment

- Also better to estimate absolute risks
  - Appropriately adjusted survival curves
  - Not only hazard ratios
Survival under 3 epoetin dosing regimes

Conclusions

1. “Current vs. never” contrast does not address a relevant clinical question
2. Observational studies are not always terribly wrong
3. Analysis of observational studies and RCTs should be the same
4. It’s all about the question
Conclusion #1: “Current vs. never” contrast not a relevant one

☐ Need to use incident (new) users
☐ New-users design is increasingly frequent in pharmacoepidemiology
☐ We discussed a generalization that allows the use of incident users with time-varying treatments
  ■ Implicit in Robins’s methods since 1986
Conclusion #2: Observational studies not always terribly wrong

- High-profile “failures” were the result of not using incident users
  - Appropriate analysis decreases the randomized-observational differences

- Good news because RCTs not feasible for many questions
  - And pragmatic RCTs are always at risk of becoming glorified observational studies
Conclusion #3: Same analysis for RCTs and observational studies

- Except for additional adjustment for baseline confounding in observational studies
  - And perhaps small trials

- Adjustment for time-varying confounding in observational studies equates adjustment for noncompliance in RCTs
Conclusion #4
It’s all about the question

- Intention-to-treat?
- Per-protocol?
- Other?

- Methods are secondary
- Emphasis on sophisticated statistical methods becomes a distraction
  - that obscures the real issues
• A young couple moves into an apartment and decides to repaper the dining room. They ask the neighbor who has a dining room the same size,
• “How many rolls of wallpaper did you buy when you papered your dining room?”
• “Seven”, he says
• So the couple buys seven rolls of expensive paper, and they start papering. When they get to the end of the fourth roll, the dining room is finished. Annoyed, they go back to the neighbor and say
• “We followed your advice, but we ended up with three extra rolls!”
• “So”, he says, “that happened to you too.”

(Cathcart and Klein. *Plato and a Platypus Walk into a Bar...*)