

# What was your question again?

---

The perils of using observational data to answer questions other than the one you might wish to ask



Miguel A. Hernán

Harvard School of Public Health

[www.hsph.harvard.edu/causal](http://www.hsph.harvard.edu/causal)

# How to estimate the effects of hypothetical interventions

---

(Hint: First specify the intervention)



Miguel A. Hernán

Harvard School of Public Health

[www.hsph.harvard.edu/causal](http://www.hsph.harvard.edu/causal)

# Setting:

---

- We want to make inferences about the causal effect
- of some treatment (or exposure)
  - hormone therapy, lifestyle
- on some outcome
  - CHD risk
  
- ❖ Observational data available on treatment, outcome, and other variables

# Causal inference from **already collected** observational data

---

- ❑ Smart study design is not an option
- ❑ We need to either quit or deal with whatever data we have
- ❑ A common situation in epidemiologic research
  - Perhaps even more common in social sciences
- ❑ How do we do this?

# Causal inference from **already collected** observational data

---

1. Formulate a reasonably well-defined causal question
  2. Propose an answer by combining
    - Available data
    - Untestable assumptions
    - Appropriate analytic method
- Often discussions about causal inference revolve exclusively around stage #2

# Stage 1: A reasonably well-defined causal question

---

- A prerequisite for meaningful causal inference
- Can be expressed in terms of
  - Hypothetical interventions
  - Counterfactual contrasts
- Choice of analytic approach follows naturally from it
  - i.e., formulation of causal question predates discussions about analytic approach

## Stage 2: Combination of data, assumptions, and analytic approach

---

- ❑ Data need to be measured with reasonable accuracy and in a population similar to the target population
- ❑ Assumptions need to be consistent with expert subject-matter knowledge
- ❑ Analytic approach needs to be appropriate to answer the causal question with the above data under the above assumptions

# Analytic approach in randomized vs. observational studies

---

- Randomized experiments (e.g. clinical trials):
  - analysis pre-specified before study is conducted
  - Described in study protocol
- Observational studies
  - Often ad hoc analysis after data have been collected and...
  - explored, massaged, or even tortured
  - Vulnerable to criticism (data dredging?)

# Observational studies analyzed like randomized experiments

---

- Specify the **causal question** of interest
- Design the protocol
  - eligibility criteria, regimes to be compared, period of follow-up, **analytic approach**, ...
- of a hypothetical randomized experiment to answer the causal question of interest
- Try to emulate such experiment with the observational **data + assumptions**

# Two examples

---

1. Relatively well-defined causal question, relatively well-known answer
  - Postmenopausal hormone therapy and risk of coronary heart disease (CHD)
  - Today's talk
2. Less well-defined causal question, answer unknown or known only qualitatively
  - Lifestyle and risk of CHD
  - Another talk

# EXAMPLE

## Hormone therapy and heart disease

---

- Relatively well-defined causal question
  - Does initiation of postmenopausal hormone therapy (estrogen plus progestin) increase the risk of CHD?
  - Can be expressed in terms of hypothetical interventions or counterfactuals
- Relatively well-known answer
  - A randomized experiment found >20% increased risk of CHD in initiators compared with noninitiators
  - The intervention became non hypothetical

# The WHI randomized trial

Manson et al, NEJM 2003

---

- A large double-blind randomized trial
  - >16,000 U.S. women aged 50-79 yrs
  - Randomly assigned to hormones or placebo
- Women followed approximately every year like in many large observational studies
  - No intervention after baseline
- Analytic approach: Intent-to-treat (ITT) analysis
  - Not all women adhered to their assigned treatment, some guessed their treatment

# WHI: ITT effect estimates

## Hazard ratio (95% CI) of CHD

---

<input type="checkbox"/>	Overall	1.23 (0.99, 1.53)
<input type="checkbox"/>	Years of follow-up	
<input type="checkbox"/>	0-2	1.51 (1.06, 2.14)
<input type="checkbox"/>	>2-5	1.31 (0.93, 1.83)
<input type="checkbox"/>	>5	0.67 (0.41, 1.09)
<input type="checkbox"/>	Years since menopause	
<input type="checkbox"/>	<10	0.89 (0.54, 1.44)
<input type="checkbox"/>	10-20	1.24 (0.86, 1.80)
<input type="checkbox"/>	>20	1.65 (1.14, 2.40)

# Before the WHI

---

- Several observational studies published in the 1980s and 1990s had apparently found the opposite result
  - lower CHD risk in users of hormone therapy compared with nonusers
- For example, in the Nurses' Health Study the CHD hazard ratio for current versus never use was **0.68** (0.55, 0.83)
  - Grodstein et al (*J Women's Health* 2006)
- Clinical recommendations were based on the estimates from observational studies

# Before the WHI

---

## **1992 American College of Obstetricians and Gynecologists**

*"Probable beneficial effect of estrogen on heart disease"*

## **1992 American College of Physicians**

*"Women who have coronary heart disease or who are at increased risk of coronary heart disease are likely to benefit from hormone therapy"*

## **1993 National Cholesterol Education Program**

*"Epidemiologic evidence for benefit of estrogen replacement therapy is especially strong for secondary prevention in women with prior CHD"*

## **1996 American Heart Association**

*"ERT does look promising as a long-term protection against heart attack"*

# After the WHI: Chain reaction

---

- ❑ There is a clear discrepancy
- ❑ Since randomized trials are the gold standard for causal inference...
- ❑ Observational studies got it wrong
- ❑ Can observational studies ever be trusted again?
  - The end of observational epidemiology?
- ❑ *Should we fund observational studies?*

# Why did observational studies get it “wrong”?

---

- Popular answer: the key **assumption** of no unmeasured confounding was violated
  - the randomized-observational discrepancy is due to insufficient adjustment for lifestyle risk factors and socioeconomic indicators
  - Corollary: causal inference from observational data is a hopeless undertaking
- Consider an alternative answer: **causal question** was not explicit
  - Observational studies implicitly asked a question different from the question explicitly asked by the randomized trial

# More formally, the different (discrete-time) hazard ratios were

---

## □ Randomized: Initiators vs. noninitiators

- $\Pr[Y_{t+1}=1|Y_t=0, A_0=1] /$   
 $\Pr[Y_{t+1}=1|Y_t=0, A_0=0]$  for  $t \geq 0$

## □ Observational: Current vs. never users

- $\Pr[Y_{t+1}=1|Y_t=0, A_t=1] /$   
 $\Pr[Y_{t+1}=1|Y_t=0, A_m=0 \text{ for } 0 \leq m \leq t]$

where

- $Y_t$ : CHD indicator at time  $t$
- $A_t$ : indicator for use of therapy at time  $t$

# Randomized experiment

---

- First state your question, then decide your analytic approach
  - Explicit causal question: what is the effect of hormone therapy **initiation** on CHD risk?
  - Analytic approach following from that question: Compare risk between women who initiate and do not initiate hormone therapy (ITT)

# Observational studies

---

- First decide your analytic approach, then try to find out the question you are answering?
  - Analytic approach: Compare risk between women who currently use therapy and those who never used it
  - Implicit causal question: what is the effect of hormone therapy **continuation** on CHD risk?

# Our strategy

---

- Use the observational data to answer same question as randomized experiment
  - Re-analyze observational studies to estimate the observational analog of the ITT effect
- Then compare both set of estimates
  
- For a detailed description see
  - Hernán et al. *Biometrics* 2005
  - Hernán et al. *Epidemiology* 2008
  - Hernán and Robins. *Epidemiology* 2008

# The Nurses' Health Study (NHS)

---

- A large longitudinal observational study
  - >120,000 women recruited in 1976
  - ~80,000 with complete 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
- 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 use of oral estrogens plus progesterone 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

---

- (WHI baseline: randomization time)
- NHS baseline:
  - Initiators: month of initiation in 2-yr period before the 1984 questionnaire
  - Non initiators: average baseline month among initiators
- 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 nonrandomized study 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
    - e.g., shorter time since menopause in NHS than in WHI
  - length of follow-up
    - longer in NHS than in WHI

# Protocol of the NHS “trial”

## Intention to treat (ITT) principle

---

- Compare the risk of CHD between women who initiated and did not initiate hormone therapy at baseline
  - Conditional on potential confounders
- Regardless of future hormone use during the follow-up
- This is the observational analog of the ITT effect

# Protocol of the NHS “trial”

## Analytic approach

---

- ❑ Same as in WHI: Cox proportional hazards model
- ❑ Covariates:
  - Indicator for hormone therapy initiation
  - Age, past hormone use, parental history of myocardial infarction before age 60, 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 and vegetable intake

# The NHS “trial”

## Non randomized after all

---

- To obtain valid effect ITT estimates in a nonrandomized trial, all baseline confounders have to be appropriately measured and adjusted for in the analysis
  - We proceeded as if this condition was at least approximately true in the NHS trial after adding the above covariates to the Cox model
  
- **Untestable assumption:** combined with **data** and **analytic approach** to answer the **causal question** of interest
  - Unnecessary assumption in truly randomized experiments

# The NHS “trials”

---

- We started the NHS “trial” during the period before the 1994 questionnaire but there is nothing special about the 1984 questionnaire
- We can start our “trial” in the period before the 1986, 1988, ... or 1998 questionnaires
  - Sequence of nested “trials”
- Or we can conduct all possible “trials,” pool the data across “trials,” and obtain an effect estimate with a narrower confidence interval
  - Need to adjust for within-subject correlation

# The NHS “trials”

---

- We started a separate NHS trial before each questionnaire  $m$ 
  - $m=0,1,\dots, 8$  representing 1984, 1986,... 1998
- Each woman may participate in a maximum of 8 trials
- For each trial,
  - follow-up started at the trial-specific baseline (as defined above) and ended at diagnosis of a CHD endpoint, death, lost to follow-up, or June 2000, whichever came first
  - Eligibility criteria applied at baseline

# Analytic approach (Nested) Cox model

---

$$\lambda_T[t|G(m) = 1, A(m), \bar{L}(m)] = \lambda_0[t] \left[ \alpha A(m) + \theta'_1 \bar{L}(m) \right]$$

## □ Notation

- $T$ : CHD-free survival time
- $G(m)$ : indicator for eligibility at  $m$
- $L(m)$ : covariates measured before  $m$

## □ PMLE, robust variance

## □ Conditional ITT hazard ratio: $\exp(\alpha)$

## □ Similar results using doubly-robust estimators from nested structural AFT model that incorporates propensity score

# More formally, the discrete-time hazard ratios are

---

## □ Initiators vs. noninitiators

- $\frac{\Pr[Y_{t+1}=1|Y_t=0, A_t=1, A_m=0 \text{ for all } 0 \leq m < t]}{\Pr[Y_{t+1}=1|Y_t=0, A_t=0, A_m=0 \text{ for all } 0 \leq m < t]}$

## □ rather than Current vs. never users

- $\frac{\Pr[Y_{t+1}=1|Y_t=0, A_t=1]}{\Pr[Y_{t+1}=1|Y_t=0, A_m=0 \text{ for all } 0 \leq m \leq t]}$

where

- $Y_t$ : CHD indicator at time  $t$
- $A_t$ : use of therapy at time  $t$

# 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

# ITT effect estimates

## WHI

## NHS

---

□ Overall	1.23 (0.99, 1.53)	1.05 (0.82, 1.34)
□ Years of follow-up		
■ 0-2	1.51 (1.06, 2.14)	1.43 (0.92, 2.23)
■ >2	1.07 (0.81, 1.41)	0.91 (0.72, 1.16)
□ Years since menopause		
■ <10	0.89 (0.54, 1.44)	0.88 (0.63, 1.21)
■ 10-20	1.24 (0.86, 1.80)	1.13 (0.85, 1.49)
■ >20	1.65 (1.14, 2.40)	--

# Conclusions

---

- No shocking discrepancy between observational and randomized studies when the same **causal question** is asked
  - though wide confidence intervals in both the WHI and the NHS
- What about the popular response? Is there any unmeasured confounding?
  - Probably, but insufficient to explain the original discrepancy

# Conclusions

---

- The choice of the **analytic method** was guided by the **question**
  - Our sequence of nested “trials” is a particular case of g-estimation of nested structural models
- Not a vindication of ITT analyses!!
  - We used the observational data to estimate ITT effect only to facilitate comparison with the trial
- ITT analyses are problematic in the presence of substantial noncompliance
  - Better to estimate adherence-adjusted effects like the effect of continuous hormone use

# What is the effect of continuous hormone therapy versus no use?

---

- A hypothetical intervention over time
  - rather than only at baseline
- We estimated this effect in both the WHI and the NHS via inverse probability weighting
  - weaker assumptions than the “current vs. never” comparison
- Again, little discrepancy
  - Hernán et al. *Epidemiology* 2008
  - Toh et al. 2009 (under review)

# Effect estimates for continuous use

## WHI

## NHS

---

□ Overall	1.60 (0.90, 2.84)	1.30 (0.76, 2.21)
□ Years of follow-up		
■ 0-2	2.45 (1.64, 3.67)	1.71 (1.03, 2.83)
■ >2	1.25 (0.81, 1.95)	1.07 (0.44, 2.63)
□ Years since menopause		
■ <10	0.75 (0.23, 2.47)	0.68 (0.24, 1.91)
■ ≥10	1.96 (1.02, 3.76)	1.57 (0.86, 2.85)

# Conclusions

---

- A clear specification of the question of interest helps
  - state our assumptions precisely
  - compare apples with apples, or oranges with oranges
    - Randomized vs. observational ITT estimates, randomized vs observational adherence-adjusted estimates
  - We can then discuss whether we like apples or oranges better