

Observational studies analyzed
like randomized experiments,
and vice versa

The case of postmenopausal hormone
therapy and heart disease

2008 Mid-Atlantic Causal Inference

Johns Hopkins Bloomberg

School of Public Health

May 19 – 20, 2008

Session organizer: Miguel A. Hernán

Overview of the session

Overview of the Women's Health Initiative

- Jacques Rossouw, NHLBI

Observational studies analyzed like randomized experiments

- Miguel Hernán, Harvard School of Public Health

10 min break

Randomized experiments analyzed like observational studies

- Darren Toh, Harvard School of Public health

Discussion

- James Robins, Harvard School of Public Health

Open discussion

Observational studies analyzed like randomized experiments

The case of postmenopausal hormone therapy and heart disease



Miguel A. Hernán

Harvard School of Public Health

www.hsph.harvard.edu/causal

Joint work with A Alonso, R Logan, F Grodstein, K Michels, M Stampfer, W Willett, J Manson, J Robins
Funded by NIH R01 grant HL080644

The issue

- Postmenopausal hormone therapy (estrogen plus progestin) and risk of coronary heart disease (CHD)
- Observational studies found a lower risk in users throughout the 1980s and 1990s
 - Nurses' Health Study (NHS), General Practice Research Database (GPRD)
 - Hazard ratios: 0.5-0.7
- A randomized experiment found a greater risk in users in 2002
 - Women's Health Initiative (WHI)
 - Hazard ratio: 1.24

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?*

Let's step back for a second

- ❑ What is the design of the WHI randomized trial for estrogen+progestin?
- ❑ How does it differ from an observational study?

The WHI trial

- A large double-blind randomized trial
 - >16,000 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
- During the “observational” follow-up, many women
 - did not adhere to their assigned treatment
 - guessed the treatment they were receiving

The WHI

- Randomized intervention at baseline
- Observational follow-up

- Longitudinal study with baseline randomization
 - Large simple trial (LST)

How to analyze a longitudinal study with baseline randomization?

Strict trialists

- Estimate intention-to-treat effect (ITT)
- Of course!

Others

- ITT effect may be biased because of incomplete adherence (noncompliance)
 - especially problematic for safety outcomes
- Estimate some sort of “adherence-adjusted” effect
 - typically, current users vs never users

WHI: ITT effect estimates Hazard ratio (95% CIs) 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)

Two possible directions for this talk

ITT estimates

- Compare estimates from WHI and observational studies *without* baseline randomization

Adherence-adjusted effect estimates

- Compare adherence-adjusted estimates from WHI and observational studies *without* baseline randomization

We take direction 1

- Next talk will take direction 2

The Nurses' Health Study (NHS)

- A longitudinal study without baseline randomization
 - >120,000 women recruited by questionnaire in 1976
 - ~80,000 with dietary 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

The Nurses' Health Study (NHS)

- Most recently published effect estimate of hormone therapy on CHD risk:

0.68 (0.55, 0.83)

- CHD Hazard ratio for current vs never users
- Grodstein et al (2006)
- Looks very different from WHI estimate
- But it is not directly comparable with WHI estimate
 - No current vs never users comparison in WHI!
 - No ITT comparison in NHS!

Our strategy

- Re-analyze the NHS like the WHI
 - **Estimate the observational analog of the ITT effect in the NHS**
 - Need to conceptualize the observational study as a sequence of trials

- Then compare the ITT estimates from the NHS and the WHI

The NHS “trial”

Eligibility criteria


□ Eligibility criteria

- women aged 50 years or more and with an intact uterus
- no past diagnosis of cancer (except non melanoma skin cancer), acute myocardial infarction, or stroke
- dietary data in 1980 (for adjustment purposes)

□ Similar to WHI criteria

The NHS “trial”

Treatment regimes

 Initiation of use of oral estrogens plus progesterone at baseline

 No hormone use at baseline

Washout interval: no hormone use in 2-yr period before baseline (additional eligibility criterion)

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

- Different
 - 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

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 in the WHI

The NHS “trial”

Analytic approach

- 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: **the key difference between nonrandomized and randomized studies**
 - Otherwise, longitudinal studies with and without baseline randomization look pretty much the same

The NHS “trials”

- There is nothing special about the period before 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 the variance of the estimate
- Eligibility criteria applied at each trial baseline

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

Analytic approach (Nested) Cox model

$$V_T \beta | G Y_m = 1, A Y_m, L Y_m = V_0 \beta \left[J A Y_m + S_1^V L Y_m \right]$$

- 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

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

The NHS trials

Trial	Questionnaire year	Participants	Initiators	CHD events	
				All	Initiators
1	1984	16,190	224	772	10
2	1986	17,147	518	671	9
3	1988	18,620	979	610	17
4	1990	19,002	1,082	528	14
5	1992	19,494	1,152	441	13
6	1994	19,954	1,344	354	19
7	1996	19,661	1,188	228	11
8	1998	18,192	771	100	5

NHS: ITT effect estimates

Hazard ratio (95% CIs) of CHD

- Overall 1.05 (0.82, 1.34) [1st 8 yrs of follow-up]

- Years of follow-up
 - 0-2 1.43 (0.92, 2.23)
 - >2 0.91 (0.72, 1.16)

- Years since menopause
 - <10 0.88 (0.63, 1.21)
 - >10 1.13 (0.85, 1.49)

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)	--

Sensitivity analyses for several analytic decisions

- ❑ Determination of month of therapy initiation
- ❑ Exclusion of women who died between start of follow-up and return of baseline questionnaire
- ❑ Confounding adjustment via propensity scores

Estimates insensitive to these decisions

Discussion: Assumption of no unmeasured confounding

- Popular theory:
 - the NHS-WHI discrepancy can be explained by insufficient adjustment for lifestyle factors and socioeconomic indicators
- In our study, adjustment for lifestyle factors and socioeconomic indicators had little effect on the estimates
- A small downward bias in our estimate is still plausible but insufficient to explain the original NHS-WHI discrepancy

Discussion

Sampling variability

- Major problem
- Overall ITT hazard ratios from the NHS and the WHI trials were estimated with similar low precision
 - width of the 95% CIs on the log scale: about 0.46 in WHI and 0.45 in NHS
- This relatively low precision precludes drawing strong conclusions from either study
 - For period-specific and age-specific estimates

Discussion

Different age distribution

- WHI participants are older on average than participants in the observational studies
- This may be important if the effect of hormones varies depending on the time between menopause and treatment initiation
 - e.g., hormones may increase the risk of CHD mostly in women at a more advanced stage of atherosclerosis

Methodological conclusions

- A direct comparison between the estimates of observational studies and randomized trials can be misleading
 - Randomized trials analyzed under ITT principle
 - Observational studies typically analyzed using the 'as treated' principle
- Fair comparison requires a more comparable analytic approach like ours
 - a particular case of Robins's g-estimation of structural nested models
- Next step: comparison of NHS-WHI adherence-adjusted effect estimates

Practical conclusions regarding the discrepancy WHI/NHS

- Under our analytic approach, small difference between randomized and observational estimates
- Consistent with
 - small amount of unmeasured confounding
 - random variability
- Had the NHS been analyzed under this approach, WHI results would not have been that surprising
 - Think how much paper we would have saved