How do we learn what works?

A two-step algorithm for causal inference from real world data

Miguel Hernán

HARVARD T.H. CHAN
SCHOOL OF PUBLIC HEALTH

The starting point:
We need to make decisions NOW

- Treat with A or with B?
- Treat now or later?
- Switch to C?
- Stop all treatment?
- etc.

- Decision making needs to be informed by causal knowledge about comparative effectiveness
  - and safety

How do we learn what works and what harms?
(How do we estimate causal effects?)

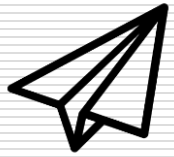
- The standard scientific answer:
  - Conduct a randomized experiment

- A relevant randomized trial would, in principle, answer each causal question about comparative effectiveness and safety
  - Interference/scaling up issues aside

But we rarely have randomized trials

expensive unethical impractical untimely






- And deferring decisions is not an option
  - no decision is a decision: “Keep status quo”
- What do we do?
  - We analyze observational data

Types of observational data

Research data

- Data collected specifically for research
  - Cohort studies, case-control studies, and other epidemiologic studies
  - Biobanks
  - Disease registries
  - …

Found data

- Data generated for non-research purposes
  - Electronic health records
  - Insurance claims databases
  - National registers
  - …

"Real world data"

We analyze observational data

because we cannot conduct a randomized trial

Observational analyses are **not** our preferred choice

- For each observational analysis for causal inference, we can imagine a hypothetical randomized trial that we would prefer to conduct
  - If only it were possible

The Target Trial

- The (hypothetical) randomized trial that we would like to conduct to answer a causal question
  - To learn what works and what harms

- A causal analysis of observational data can be viewed as an attempt to emulate some target trial

The Target Trial

- Suggested more or less explicitly by many authors
  - Dorn (1953), Cochran, Rubin, Feinstein, Dawid…
  - for simple settings with a time-fixed treatment and a single eligibility point

- Explicit generalization to time-varying treatments and multiple eligibility points
  - Robins (1986)
  - Hernán, Robins. *Am J Epidemiol* 2016

The Target Trial concept leads to a simple algorithm for causal inference

1. Ask a causal question (point at the Target)
   - Specify the protocol of the Target Trial
2. Answer the causal question (shoot the Target)
   - Option A
     - Conduct the Target Trial
   - Option B
     - Use observational data to **explicitly** emulate the Target Trial
     - Apply appropriate causal inference analytics

Step 1
Specify Target Trial protocol

- Eligibility criteria
- Treatment strategies
- Randomized assignment
- Start/End follow-up
- Outcomes
- Causal contrast(s) of interest
- Analysis plan

Step 2
Emulate Target Trial protocol

- Eligibility criteria
- Treatment strategies
- Randomized assignment
- Start/End follow-up
- Outcomes
- Causal contrast(s) of interest
- Analysis plan

Ok, so why is this a big deal?

- Why do we need to explicitly need to emulate a target trial when using observational data to learn what works?
- What happens if we just analyze the data as usual?
  - That is, if we compare “exposed” vs. “unexposed” and adjust for covariates?

- Let’s see an example

EXAMPLE #1
Postmenopausal hormone therapy and heart disease

- Observational epidemiologic studies
  - >30% **lower risk** in current users vs. never users
    - e.g., hazard ratio: 0.68 in Nurses' Health Study
      - Grodstein et al. *J Women's Health* 2006
- Randomized trial
  - >20% **higher risk** in initiators vs. noninitiators
    - hazard ratio: 1.24 in Women's Health Initiative
      - Manson et al. *New England J Med* 2003

Shocking discrepancy!

The randomized trial
Women's Health Initiative (WHI)

- Double-blind
- Placebo-controlled
- Large
  - >16,000 U.S. women aged 50-79 yrs
- Randomly assigned to
  - estrogen plus progestin therapy
  - placebo
- Women followed approximately every year
  - for a maximum of 8 years

WHI: Effect estimates

Intention-to-treat hazard ratio (95% CI) of coronary heart disase

- 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)

This hazard ratio can be fully explained by selection bias even if no woman benefits from hormone therapy

(Stensrud et al. *Epidemiology* 2017)

Why did observational studies get it "wrong"?

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

- An alternative theory
  - The observational studies were not emulating a target trial

WHI randomized trial compared women who **initiated** therapy with women who did not

- Design
  - Women randomly assigned to initiation of hormone therapy or placebo
  - Almost all women assigned to initiation received at least a dose, that is, they are classified as initiators
- Analysis
  - Compared risk between initiators (**incident** users) and noninitiators of hormone therapy

- This trial informs decisions about therapy initiation

Observational studies compared women **currently using** therapy with women who did not use it

- Design
  - Women were asked about therapy use
  - They were classified as current, past, or never users
- Analysis
  - Compared risk between current (**prevalent**) users and never users of hormone therapy
    - Was the estimate different from that of the WHI trial?
- What decision does this design/analysis inform?
  - What is the target trial?

What if we re-analyze the observational data…

… to explicitly emulate a target trial as close as possible to the WHI trial?

- Causal inference algorithm
  - Step 1: Specify the protocol of a target trial of hormone therapy and coronary heart disease
  - Step 2: Emulate it
    - Hernán et al. *Biometrics* 2005; 61(4):922–930
    - Hernán et al. *Epidemiology* 2008; 19(6):766-779

Step 1
Specify Target Trial protocol

- Eligibility criteria
- Treatment strategies
- Randomized assignment
- Start/End follow-up
- Outcomes
- Causal contrast(s) of interest
- Analysis plan

Step 2
Emulate Target Trial protocol

- Eligibility criteria
- Treatment strategies
- Randomized assignment
- Start/End follow-up
- Outcomes
- Causal contrast(s) of interest
- Analysis plan

Summary of Protocol of Target trial: Hormone therapy and coronary heart disease

Eligibility criteria	Postmenopausal women with no history of cancer and other diseases, and no use of hormone therapy in the last 2 years.
Treatment strategies	1. Initiate estrogen plus progestin hormone therapy at baseline and remain on it during the follow-up, unless deep vein thrombosis, pulmonary embolism, myocardial infarction, or cancer are diagnosed 2. Refrain from taking hormone therapy during the follow-up
Assignment procedures	Participants will be randomly assigned to either strategy at baseline, and will be aware of the strategy they have been assigned to.
Follow-up period	Starts at randomization and ends at coronary heart disease diagnosis, death, loss to follow-up, or June 2000, whichever occurs earlier.
Outcome	Coronary heart disease diagnosed by a cardiologist
Causal contrasts	Intention-to-treat effect, per-protocol effect
Analysis plan	Intention-to-treat analysis, non-naïve per-protocol analysis

Observational data for emulation: The Nurses' Health Study

- Epidemiologic follow-up (cohort) study
- ~80,000 women with full data in 1980
- Information updated by questionnaire every two years
  - Use of hormone therapy
  - Diagnosis of coronary heart disease (confirmed by physician)
  - Medical diagnoses
  - Lifestyle data: diet, exercise, smoking…
  - Other risk factors for coronary heart disease

Emulation: Intention-to-treat analysis

- Compare CHD incidence between initiators and noninitiators of hormone therapy at baseline
  - Regardless of future use during the follow-up

- Fit a Cox model (like the WHI) with covariates
  - Age, past hormone use, parental history of myocardial infarction before age 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

Emulation summary

- We used the observational data to emulate a target trial with similar eligibility criteria, treatment arms, outcome, causal contrast, and analysis plan as the WHI randomized trial

- Some differences
  - Not blinded
  - Not placebo-controlled
  - Shorter average time since menopause than WHI
  - Longer follow-up than WHI

Effect estimates: hazard ratios (95% CIs)

	Randomized Women's Health Initiative	Observational Nurses' Health Study
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)	--

When the target trial is explicitly emulated, then the same **causal question** is asked

- No shocking observational-randomized discrepancies
  - though wide confidence intervals in both studies

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

Epidemiologic studies may be adequate to emulate target trials

- If high-quality observational data on treatment, outcome, and confounders are available
  - e.g., the Nurses' Health Study
- But most observational research relies on real world data
- Can emulation of a target trial work with large databases of real world data?
  - Let's see some examples

Examples of Target Trial emulation using different types of observational data

1. Hormone therapy and heart disease
   - Research data: Epidemiologic study
2. Statins and mortality in cancer patients
   - Research data: Cancer registry
   - Real world data: Insurance claims
3. Screening colonoscopy and cancer
   - Real world data: Insurance claims
4. Statins and coronary heart disease
5. Antiretrovirals and mortality in HIV-positive individuals
   - Real world data: Electronic health records
6. Epoetin therapy and mortality in dialysis patients
   - Real world data: Insurance claims + supplementary data

EXAMPLE #2
Statins and mortality in cancer patients

- Statins are drugs that lower LDL-cholesterol
- In observational studies of cancer patients, statin use is associated with 30% lower mortality
  - Statins inhibit cancer growth?
- However, those studies did not attempt to explicitly emulate a target trial
- We did
  - Emilsson et al. *JAMA Oncology* 2017; (in press)

Summary of Protocol of Target trial: Statin therapy and mortality in cancer patients

Eligibility criteria	Individuals with Stage I-III colorectal, breast, prostate, and bladder cancer diagnosed at age 66 years or older, enrolled in Medicare parts A-B-D, and who did not receive a statin prescription in the previous 6 months.
Treatment strategies	1. Initiate statin therapy within 6 months of cancer diagnosis; discontinuation at any time that is clinically indicated 2. Refrain from using statin therapy during the follow-up
Assignment procedures	Participants will be randomly assigned to either strategy at baseline, and will be aware of the strategy they have been assigned to.
Follow-up period	Starts at randomization and ends at death, loss to follow-up, or December 2011, whichever occurs earlier.
Outcome	Cancer-specific mortality and all-cause mortality
Causal contrasts	Intention-to-treat effect, per-protocol effect
Analysis plan	Intention-to-treat analysis, non-naïve per-protocol analysis

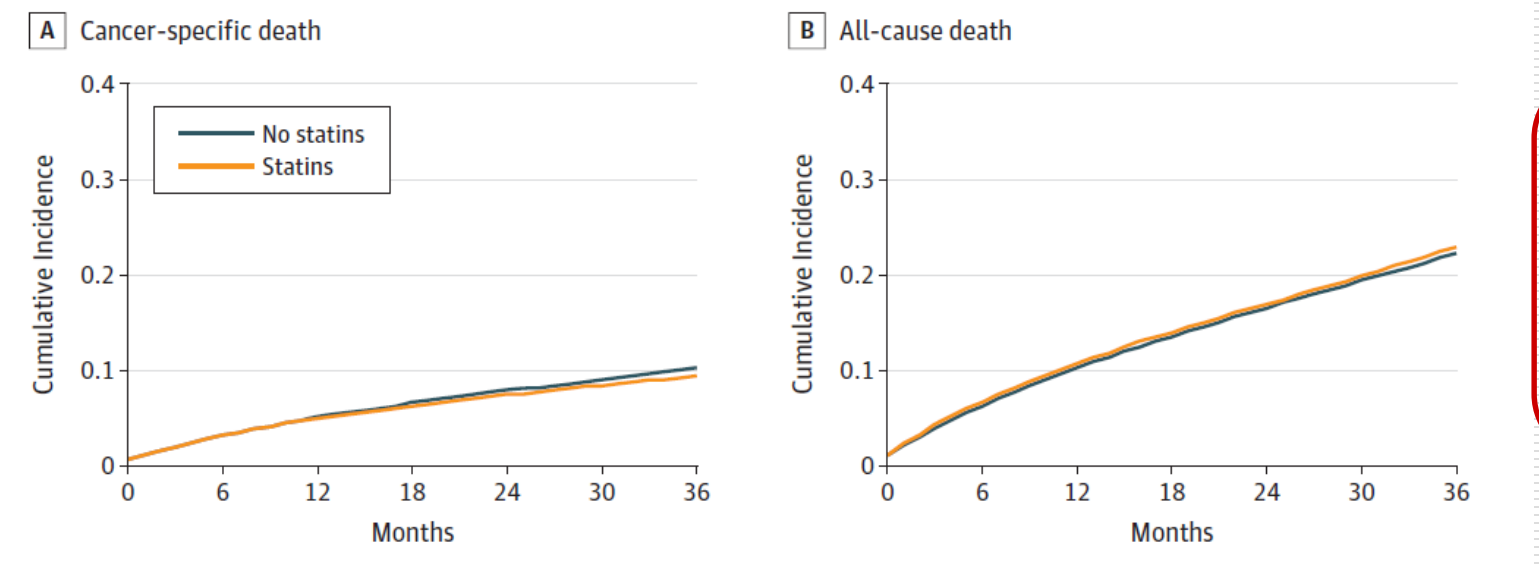
Observational data for emulation: SEER-Medicare

- SEER
  - cancer registries in 12 U.S. states
  - detailed information about cancer diagnosis
- U.S. Medicare
  - health insurance program for people 65 years or older (and others)
  - database includes insurance claims for all services provided, including statins, and death

- SEER-Medicare is the linkage of both

SEER-Medicare emulation: Hazard ratio estimates for statin vs. no statin initiation

- Cancer-specific mortality: 1.00 (0.88, 1.15)
- All-cause mortality: 1.07 (0.93, .21)



No beneficial effect of statins?
What about previous observational studies?

Emulating time zero (start of follow-up) is crucial to learn what works

- Criticisms of observational analyses often focus on residual confounding
  - failure to emulate randomization because of insufficient data on confounders
  - Hard to fix
- But many observational analyses have a more fundamental problem
  - Failure to choose time zero
  - Easy to fix

Time zero of follow-up in the Target Trial

- The time when 3 things happen
  - eligibility criteria are met
  - treatment strategies are assigned
  - study outcomes begin to be counted
- The same applies to observational analyses that emulate a target trial
- Misalignment of eligibility criteria and treatment assignment leads to selection bias / immortal time bias
  - Hernán et al. *J Clin Epidemiol* 2016; 79:70-75.

Emulation of time zero is not straightforward when there are multiple eligibility times

- In Example #2 (Statins in cancer patients), eligibility criteria are met as a single time
  - Cancer diagnosis
  - That's time zero
- In Example #1 (Hormone therapy), eligibility criteria may be met at different times
  - While a postmenopausal woman has no history of chronic disease and no hormone therapy use in the previous 2 years
  - What's time zero?

Examples of Target Trial emulation using different types of observational data

1. Hormone therapy and heart disease
2. Statins and mortality in cancer patients
3. Screening colonoscopy and cancer
4. Statins and coronary heart disease
5. Antiretrovirals and mortality in HIV-positive individuals
6. Epoetin therapy and mortality in dialysis patients

Research data: Epidemiologic study

Research data: Cancer registry
Real world data: Insurance claims

Real world data: Insurance claims

EXAMPLE #3
Screening colonoscopy and colorectal cancer

- Colonoscopy screening recommended at age 50 in US
  - but its effectiveness never proven in randomized trials
  - 3 ongoing trials; results in 2025
- Very hard to conduct randomized trials
  - 10-15 years of follow-up are needed
  - >50,000 individuals needed
  - trials do not include older patients
- Need observational data to emulate a target trial
  - Garcia-Albeniz et al. *Ann Int Med* 2017; 166(1):18-26

Summary of Protocol of Target trial
Screening colonoscopy and colorectal cancer

Eligibility criteria	Individuals aged 70–74 in 2004-2012 with no history of inflammatory bowel disease, adenoma, colectomy, and screening in the last 5 years; no gastrointestinal symptoms in last 6 months; continuous enrolment in Medicare for the last 5 years; at least 2 of the 3 preventive services offered yearly by Medicare (wellness visit, influenza vaccine, and breast or prostate cancer screening) in the previous 2 years
Treatment strategies	1. Screening colonoscopy at baseline 2. No screening colonoscopy at baseline
Assignment procedures	Participants will be randomly assigned to either strategy at baseline, and will be aware of the strategy they have been assigned to.
Follow-up period	Starts at randomization and ends at diagnosis of colorectal cancer, death, loss to follow-up, or January 2007, whichever occurs earlier.
Outcome	Colorectal cancer
Causal contrasts	Intention-to-treat effect, per-protocol effect
Analysis plan	Intention-to-treat analysis, non-naïve per-protocol analysis

U.S. Medicare

- Federal health insurance program for people 65 years or older, with disabilities or with ESRD
  - About 50 million enrollees per year
- Medicare claims dataset (20% random subsample) available for research purposes, years 1999-2012.
  - outpatient and inpatient services
  - doctor services
  - drug prescriptions

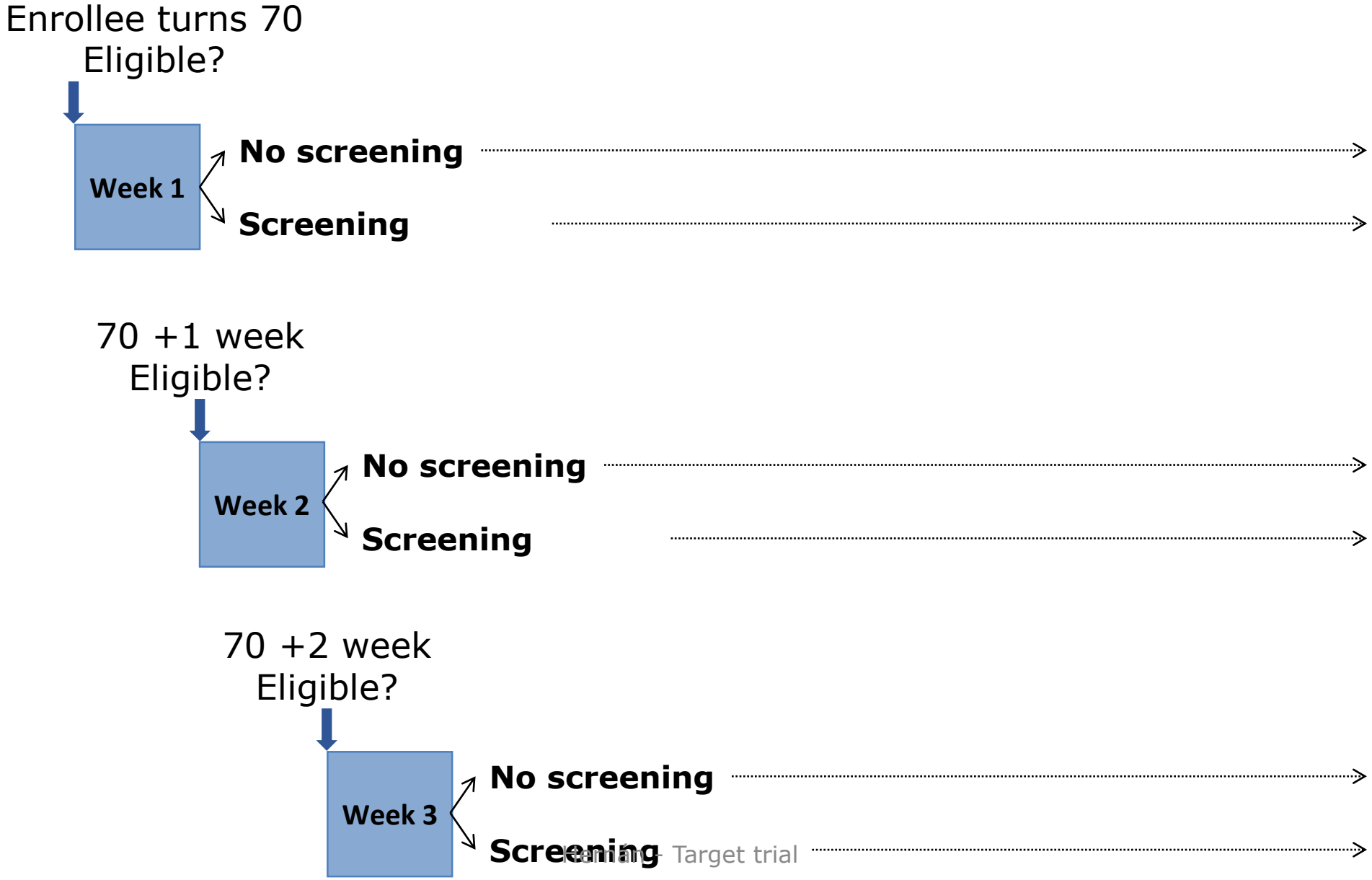
- Medicare reimburses screening colonoscopy since July 2001
  - for people at average risk for colorectal cancer without age limit

Choosing Time Zero

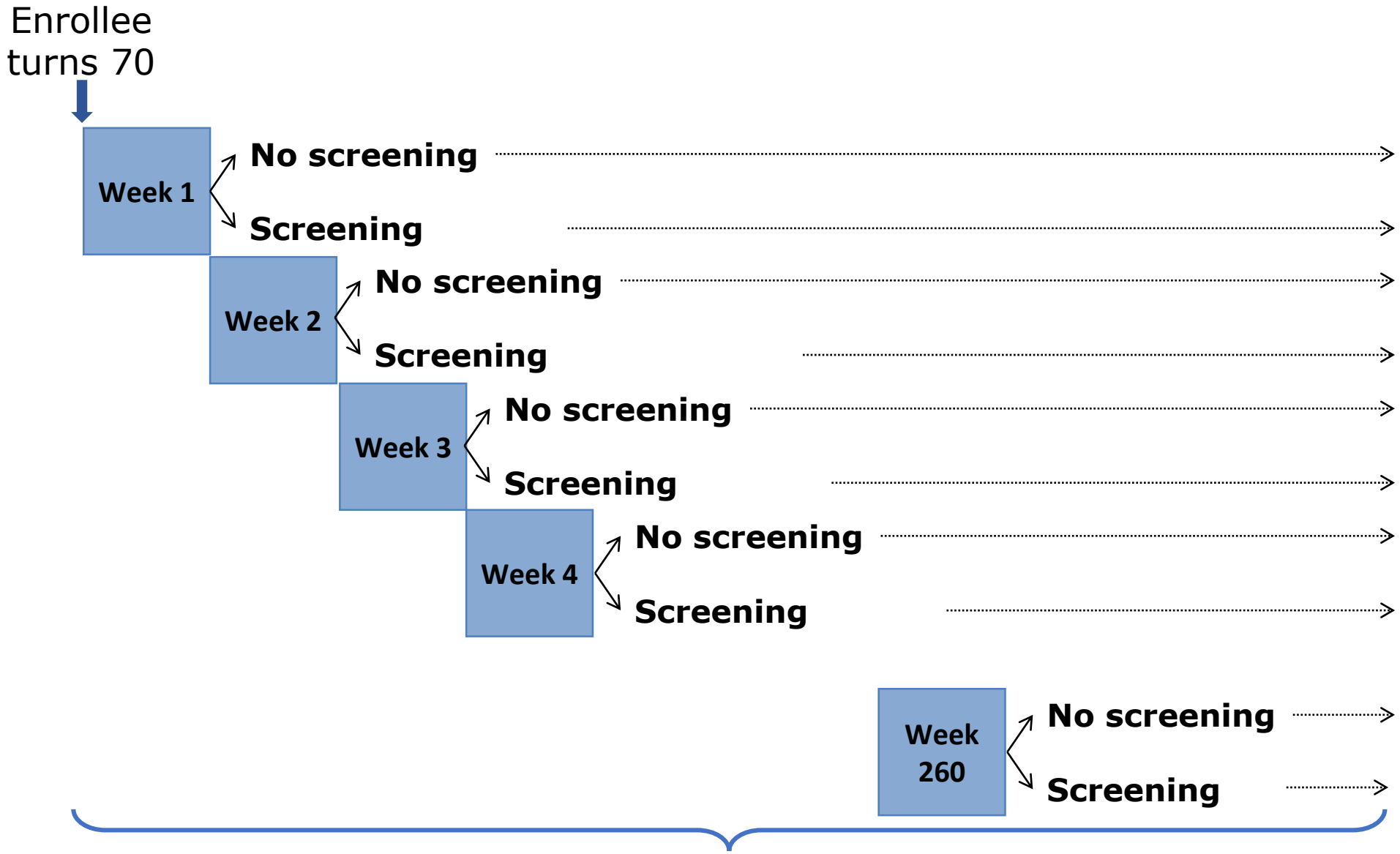
- In true trial, time zero is the time when enrollees meet the eligibility criteria
  - Outcomes are counted after time zero
- In our emulation, Medicare enrollees can meet eligibility criteria at multiple times
  - every day since they turn 70 until 74
- Two unbiased choices to choose time zero:
  - A single eligible time, e.g., the first eligible time or a random eligible time
  - **Every eligible time**: emulate a new trial starting at each eligible time

Sequential emulation

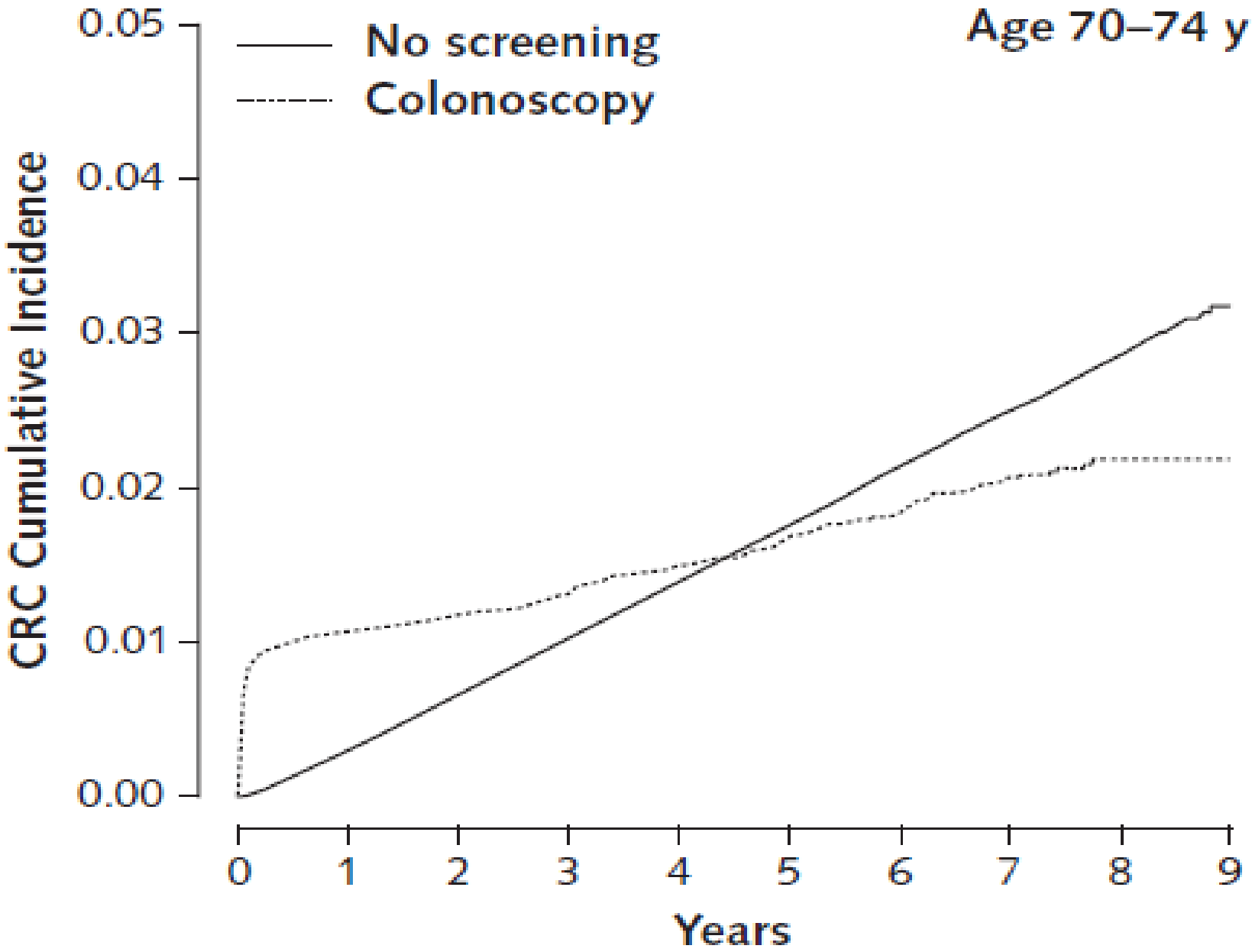
Sequential emulation



Sequential emulation



Pooling of all the person-months from the 260 "trials"


No screening
Colonoscopy
Age 70–74 y
CRC Cumulative Incidence
0.05
0.04
0.03
0.02
0.01
0.00
0
1
2
3
4
5
6
7
8
9
Years

Why make it so complicated?

Garcia-Albeniz et al. *Eur J Epid* 2017

- Consider 3 alternative observational analyses of the same data

1. Same as we did, but choosing a single time zero for each individual (e.g., the first eligible one)
   - Unbiased
   - Less efficient (wider 95% CIs)

2. Redefine no-colonoscopy group: no colonoscopy during the follow-up

- Assign individuals who
  - received a colonoscopy while meeting the eligibility criteria to the colonoscopy strategy (time zero = time of colonoscopy)
  - did not receive a colonoscopy throughout the entire study period to the no-colonoscopy group (time zero = first eligible time)
- Biased because most CRCs are eventually diagnosed via colonoscopy
  - individuals in the no-screening strategy group have little opportunity to have a CRC diagnosed
  - similar to naïve per-protocol analyses in randomized trials

3. Select arbitrary time zero (say, January 1 2004) and look back

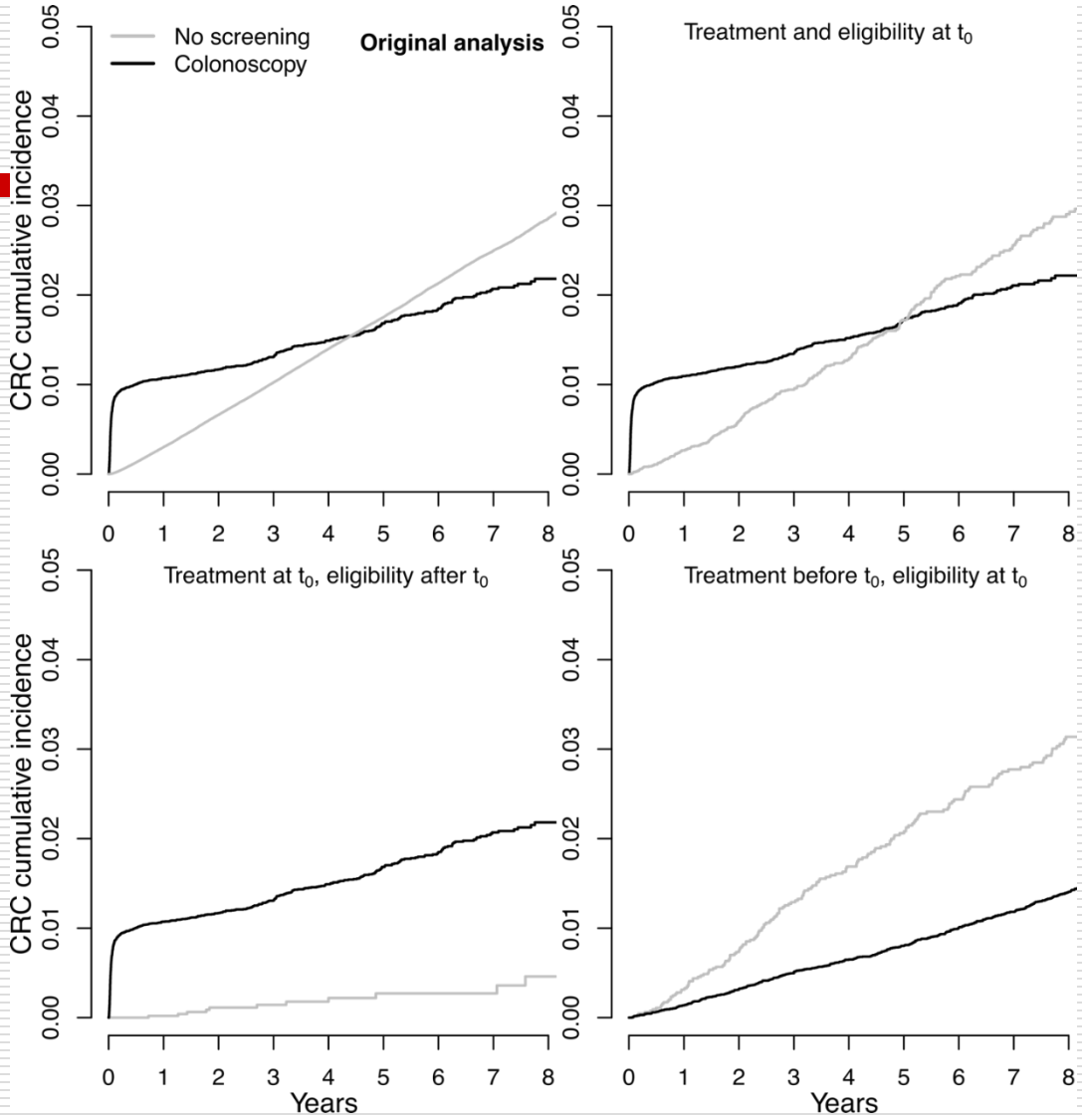
- Assign eligible individuals to
  - the colonoscopy strategy if they received a colonoscopy in the previous five years
  - the no-screening strategy otherwise.
- Bias because colonoscopies performed before assessing eligibility may affect eligibility
  - a colonoscopy that detects CRC or precursor lesions in the previous five years will result in the individual being excluded from the analysis
  - similar to approach that created confusion about the effect of postmenopausal hormone therapy in observational studies

Colonoscopy screening and 8-year CRC risk under four observational analyses

Treatment Assigned	Eligibility determined	Indiv. used multiple times	Strategy	N	CRC cases	CRC risk, % (95% CI)	Difference, % (95% CI)
At time zero	At time zero	Yes	No screening	1,762,816	21,954	2.8 (2.7, 2.8)	Ref.
			Screening	46,872	685	2.1 (2.0, 2.3)	-0.63 (-0.83, -0.43)
At time zero	At time zero	No	No screening	72,249	1086	2.8 (2.5, 3.2)	Ref.
			Screening	46,872	685	2.2 (2.0, 2.4)	-0.67 (-1.03, -0.28)
At time zero	After time zero	No	No screening	6,241	11	0.4 (0.2, 0.7)	Ref.
			Screening	46,872	685	2.1 (1.9, 2.3)	1.7 (1.4, 2.1)
Before time zero	At time zero	No	No screening	6,507	178	3.1 (2.7, 3.6)	Ref.
			Screening	37,844	492	1.4 (1.3, 1.5)	-1.8 (-2.2, -1.3)

Basic principle of trial design

- Treatment assignment and the determination of eligibility occur simultaneously at time zero
  - Observational analyses that violated this principle yielded implausible estimates



2 key components of the emulation of the target trial

1. Randomization
   - Emulation requires adjustment for confounding
2. Specification of time zero
   - Emulation requires that time zero is synchronized with determination of eligibility and assignment of treatment strategies

- Lack of randomization is usually blamed for the failings of observational analyses, but…
- We have seen that incorrect specification of time zero is often the actual culprit

Examples of Target Trial emulation using different types of observational data

1. Hormone therapy and heart disease
2. Statins and mortality in cancer patients
3. Screening colonoscopy and cancer
4. Statins and coronary heart disease
5. Antiretrovirals and mortality in HIV-positive individuals
6. Epoetin therapy and mortality in dialysis patients

Research data: Epidemiologic study

Research data: Cancer registry
Real world data: Insurance claims

Real world data: Insurance claims

Real world data: Electronic health records

EXAMPLE #4
Statins and coronary heart disease

- Randomized trials have shown that statin therapy reduces risk of coronary heart disease
- In the real world, statins are prescribed to individuals with risk factors for coronary heart disease
  - Extreme example of confounding
- Good example to test the limits of observational data (electronic health records) to emulate a target trial
  - Danaei et al. *Stat Methods Med Research* 2009; 18(1):27-52

Target trial: Statin therapy and coronary heart disease
Protocol summary

Eligibility criteria	Individuals aged 55–84 in the years 2000-2006 with no prior history of CHD, stroke, peripheral vascular disease, heart failure, cancer, schizophrenia or dementia, no symptoms of subclinical CHD, and no use of statin therapy in the last 2 years.
Treatment strategies	1. Initiate statin therapy at baseline and remain on it during the follow-up, unless contraindications arise 2. Refrain from taking statin therapy during the follow-up
Assignment procedures	Participants will be randomly assigned to either strategy at baseline, and will be aware of the strategy they have been assigned to.
Follow-up period	Starts at randomization and ends at diagnosis of coronary heart disease, death, loss to follow-up, or January 2007, whichever occurs earlier.
Outcome	Coronary heart disease
Causal contrasts	Intention-to-treat effect, per-protocol effect
Analysis plan	Intention-to-treat analysis, non-naïve per-protocol analysis

Observational data
The Health Improvement Network

- THIN is a database of electronic medical records
  - 6.2 million individuals from 350 general practices in the UK (2009)
- For each individual
  - demographic and socioeconomic characteristics
  - symptoms, signs and diagnoses, referrals, laboratory test results
  - some lifestyle information
  - vital status and cause of death data

Hazard ratio (95% CI) of CHD
THIN trials 2000-2006

	Intention-to-treat analysis
Unique cases	635
Unique persons	74,806
Cases	6,335
Person-trials	844,800
Adjusted for age and sex	1.29 (1.06, 1.56)
Adjusted for all covariates	**0.89 (0.73, 1.09)**

What if we had compared prevalent 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

These examples show that successful emulation of a Target Trial requires

- High-quality data on treatment, outcome, and confounders
  - If possible, assessment of data accuracy
    - Validation studies to quantify misclassification
    - Internal consistency checks to detect problems
    - Cross-datasets comparisons to flag coding differences
- Knowledgeable users of the data
  - Time-varying clinical workflows, idiosyncratic coding practices, software versions…
    - e.g., what does a "coronary heart disease" code mean? Maybe used when a physician suspected the diagnosis and ordered a test?

The target trial is typically a compromise

- between the ideal trial we would really like to conduct and the trial we may reasonably emulate using the available data

- The 2-step algorithm is typically iterative
  - Specifying the protocol of the target trial requires detailed knowledge of the database
  - The target trial approach allows you to systematically articulate the tradeoffs that you are willing to accept
    - regarding eligibility criteria, treatment strategies, outcomes

Advantage of the target trial approach

- Provides ready access to the application of formal **counterfactual theory** for causal inference
  - without the need for technical jargon,
- Establishes a link between methods for the analysis and reporting of randomized trials and observational studies
  - Observational studies analyzed like randomized trials, and vice versa

Every time someone presents observational estimates to estimate causal effects, ASK

"What is the target trial?"

- If they look puzzled, help them specify the target trial
- If no target trial can be identified, ask them to start over

Final thought

- As of 2017, the 2-step algorithm for causal inference cannot be fully automated

- Because the design and emulation of the target trial requires expert knowledge
  - Not yet incorporated into AIs
  - Blind analysis leads to bias

Thank you

- For more info
  - Twitter: @_MiguelHernan
  - www.hsph.harvard.edu/miguel-hernan/

- Causal Inference book
  - www.facebook.com/causalinference