How do we learn what works?
A two-step algorithm for causal inference from real world data

Miguel Hernán

DEPARTMENTS OF EPIDEMIOLOGY
AND BIOSTATISTICS

HARVARD T.H. CHAN
SCHOOL OF PUBLIC HEALTH
Harvard and Cambridge

- John Harvard statue
  Harvard University
  (Cambridge, Mass.)

- John Harvard window
  Emmanuel College
  (Cambridge, UK)
Peter Armitage CBE

www.york.ac.uk/depts/maths/histstat/people/
How do we learn what works?
A two-step algorithm for causal inference from real world data

Miguel Hernán
DEPARTMENTS OF EPIDEMIOLOGY AND BIOSTATISTICS

HARVARD T.H. CHAN SCHOOL OF PUBLIC HEALTH
Why do we want to know “what works”?
Because decisions must be made NOW

- For clinical practice
  - Treat with A or with B?
  - Treat now or later?
  - Treat all individuals?
  - Stop all treatment?

- For public health
  - Implement a screening program?
  - At what age?
  - With what frequency?
  - Until what age?

- 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”
“Routinely collected 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
  - If we cannot translate our causal question into a target trial, then the question is not well-defined
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)
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
<table>
<thead>
<tr>
<th>Step 1</th>
<th>Specify Target Trial protocol</th>
</tr>
</thead>
<tbody>
<tr>
<td>☐ Eligibility criteria</td>
<td></td>
</tr>
<tr>
<td>☐ Treatment strategies</td>
<td></td>
</tr>
<tr>
<td>☐ Randomized assignment</td>
<td></td>
</tr>
<tr>
<td>☐ Start/End follow-up</td>
<td></td>
</tr>
<tr>
<td>☐ Outcomes</td>
<td></td>
</tr>
<tr>
<td>☐ Causal contrast</td>
<td></td>
</tr>
<tr>
<td>☐ Analysis plan</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Step 2</th>
<th>Emulate Target Trial protocol</th>
</tr>
</thead>
<tbody>
<tr>
<td>☐ Eligibility criteria</td>
<td></td>
</tr>
<tr>
<td>☐ Treatment strategies</td>
<td></td>
</tr>
<tr>
<td>☐ Randomized assignment</td>
<td></td>
</tr>
<tr>
<td>☐ Start/End follow-up</td>
<td></td>
</tr>
<tr>
<td>☐ Outcomes</td>
<td></td>
</tr>
<tr>
<td>☐ Causal contrast</td>
<td></td>
</tr>
<tr>
<td>☐ Analysis plan</td>
<td></td>
</tr>
</tbody>
</table>
Ok, so why is this a big deal?

- Why do we need to explicitly emulate a target trial when using observational data to learn what works?

- Because not doing so leads to bias
  - Deviations from the target trial are the source of bias in observational analysis

- Let’s see an example
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

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
## Effect estimates from the randomized trial

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

<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><strong>Years of follow-up</strong></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><strong>Years since menopause</strong></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>

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
The 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 randomized trial?

- Causal inference algorithm
  - Step 1: Specify the protocol of a target trial of hormone therapy and coronary heart disease
  - Step 2: Emulate it
### Abbreviated Target Trial Protocol: 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
Important
Target trial must be a pragmatic trial

- Observational data cannot be used to emulate
  - a placebo-controlled trial
    - at most a trial with a “usual care” group
  - a trial with blind design
    - individuals are generally aware of the treatment they receive
  - treatment strategies that do not exist in the real world
  - enforcement of adherence to the protocol
  - tight monitoring that doesn’t happen in the real world
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

- Eligibility criteria
  - Analysis restricted to women who met the eligibility criteria of the target trial

- Treatment strategies
  1) Initiation and continued use of oral estrogens plus progesterone therapy during the follow-up
  2) No hormone therapy use during the follow-up

- Outcome
  - a diagnosis of coronary heart disease during the follow-up
Emulation: Randomized assignment

☐ This is what “adjustment for confounding” means

☐ If insufficient data on confounders, then emulation of random assignment fails
  ■ Confounding bias

☐ Need to adjust for baseline covariates
  ■ via matching, stratification or regression, standardization or inverse probability (IP) weighting, g-estimation...
Emulation: Causal contrast

- Intention-to-treat effect
  - The effect of assignment to therapy vs. no therapy
  - regardless of actual use

- Since the dataset doesn’t include prescription dates, we estimate the effect of initiation of therapy
  - Analogous to a modified intention-to-treat approach in a trial
    - Including only those who take at least one dose of treatment
Emulation: “Intention-to-treat” analysis

- Compare risk between initiators and noninitiators of therapy at baseline
  - regardless of use during the follow-up
- Fit a Cox model with an indicator for treatment initiation + confounders
  - 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, smoking, aspirin use, alcohol intake, physical activity, diet score, multivitamin use, fruit/vegetable intake
Emulation summary

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

<table>
<thead>
<tr>
<th></th>
<th>Randomized Women’s Health Initiative</th>
<th>Observational Nurses’ Health Study</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall</td>
<td>1.23 (0.99, 1.53)</td>
<td>1.05 (0.82, 1.34)</td>
</tr>
<tr>
<td>Years of follow-up</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>Years since menopause</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 the target trial is explicitly emulated, then a similar **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
What do we learn from this example?

☐ If we have an epidemiologic study
   ■ with high-quality data on treatment, outcome, and confounders (e.g., the Nurses’ Health Study)

☐ We have a fighting chance of correctly emulating a target trial

☐ But what if only have real world data?
   ■ Let’s see some examples
EXAMPLE #2
Do statins lower mortality in cancer patients?

☐ Statins are drugs that lower LDL-cholesterol
☐ Do statins inhibit cancer growth?
☐ In observational studies that did not emulate a target trial:
  ■ statin use was associated with 30% lower mortality in cancer patients
☐ What if we explicitly emulate a target trial?
  ☐ Emilsson et al. *JAMA Oncology* 2018
<table>
<thead>
<tr>
<th><strong>Summary of Protocol of Target trial:</strong> Statin therapy and mortality in cancer patients</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Eligibility criteria</strong></td>
</tr>
<tr>
<td><strong>Treatment strategies</strong></td>
</tr>
<tr>
<td><strong>Assignment procedures</strong></td>
</tr>
<tr>
<td><strong>Follow-up period</strong></td>
</tr>
<tr>
<td><strong>Outcome</strong></td>
</tr>
<tr>
<td><strong>Causal contrasts</strong></td>
</tr>
<tr>
<td><strong>Analysis plan</strong></td>
</tr>
</tbody>
</table>
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
Observational emulation: Hazard ratios for statin vs. no statin initiation

- Cancer-specific mortality: 1.00 (0.88, 1.15)
- All-cause mortality: 1.07 (0.93, 0.21)

No beneficial effect of statins? What about previous observational studies?
Selection bias in some observational studies

- Statin users at baseline vs. nonusers at baseline
  - Sounds familiar? No emulation of target trial

<table>
<thead>
<tr>
<th></th>
<th>These studies</th>
<th>When we do that</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Cancer-specific</strong></td>
<td>0.77 (0.64, 0.89)</td>
<td>0.83 (0.76, 0.91)</td>
</tr>
<tr>
<td><strong>All-cause</strong></td>
<td>0.78 (0.67, 0.90)</td>
<td>0.83 (0.79, 0.87)</td>
</tr>
</tbody>
</table>

**Mortality hazard ratio (95% CI)**
Immortal time bias in some observational studies

- Statin users at some point during the follow-up vs. nonusers during the follow-up
  - If you live longer, you are more likely to use statins

<table>
<thead>
<tr>
<th></th>
<th>These studies</th>
<th>When we do that</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Cancer-specific</strong></td>
<td>0.35 (0.27, 0.44)</td>
<td>0.31 (0.24, 0.40)</td>
</tr>
<tr>
<td><strong>All-cause</strong></td>
<td>0.39 (0.33, 0.45)</td>
<td>0.57 (0.51, 0.63)</td>
</tr>
</tbody>
</table>

Mortality hazard ratio (95% CI)
EXAMPLE #3
Do statins prevent cancer?

- Observational studies reported an association between statins and lower cancer risk
  - some studies found an implausible 50-65% lower risk
- Meta-analyses of randomized trials: No effect
  - Confounding bias due to lack of randomization?
  - Unlikely: cancer was not an intended effect of treatment
- What if we explicitly emulate a target trial?
  - Dickerman et al. *Nature Medicine* 2019
| Summary of Protocol of Target trial  
| Statins and cancer |
|---------------------|-------------------------------------------------|
| **Eligibility criteria** | Individuals aged ≥30 in January 1998-February 2016 with no history of cancer; no statin use in previous year; no statin contraindication (hepatic impairment, myopathy) LDL cholesterol <5 mml/L; at least 1 year of up-to-standard data in a CPRD practice. |
| **Treatment strategies** | 1. Initiation of any statin therapy at baseline and continuation over follow-up until the development of a contraindication  
2. No initiation of statin therapy over follow-up until the development of an indication |
| **Assignment procedures** | Participants are randomly assigned to either strategy at baseline, and are aware of the strategy they have been assigned to. |
| **Follow-up period** | Starts at randomization and ends at diagnosis of cancer, death, loss to follow-up, or February 2016, whichever occurs earlier. |
| **Outcome** | Total cancer and 7 site-specific cancers |
| **Causal contrasts** | Intention-to-treat effect, per-protocol effect |
| **Analysis plan** | Intention-to-treat analysis, non-naïve per-protocol analysis |
Observational data for emulation: CALIBER resource

- CALIBER is used to access a database (CPRD) of primary care electronic health records
  - approximately 7% of the UK population

- For each individual
  - demographics, symptoms, diagnoses
  - examination findings, lab results...
  - prescriptions and procedures
  - some lifestyle information
  - vital status
Observational emulation:
Hazard ratio estimates for statin vs. no statin

☐ Total cancer: 1.02 (0.99, 1.05)

☐ Breast cancer: 1.00 (0.92, 1.09)
☐ Colorectal: 1.04 (0.95, 1.13)
☐ Lung: 1.08 (0.99, 1.17)
☐ Prostate: 1.02 (0.95, 1.09)
☐ ...

☐ these are intention-to-treat HRs, per-protocol HRs are similar
Observational emulation:
Survival estimates for statin vs. no statin

10-year survival difference: 
-0.5% (95% CI: -1.0%, 0.0%)

No beneficial effect of statins?
What about previous observational studies?
Previous study:
Odds ratio of lung cancer: 0.23 (0.20, 0.26) for long-term users (>4 years) vs nonusers

- Two key deviations from the target trial:
  1. included prevalent users at baseline
  2. using postbaseline information (observed duration of statin therapy) to assign baseline treatment status

- When we did this in our data, the hazard ratio was:
  - 0.23 (0.22, 0.24) for total cancer
  - 0.27 (0.25, 0.29) for lung cancer
What was the original problem in these 3 examples of observational analyses?

☐ Was it lack of randomization?

NO!
Interesting state of affairs

- The usual criticism of observational analyses is lack of randomization
  - Failure to emulate randomization because of insufficient data on confounders (residual confounding)
  - Hard to fix
- Yet mounting evidence suggests another problem
  - Failure to choose a correct time zero
  - Easy to fix
Step 1
Specify Target Trial protocol

- Eligibility criteria
- Treatment strategies
- Randomized assignment
- Start/End follow-up
- Outcomes
- Causal contrast
- Analysis plan

Handling time zero correctly: The low-hanging fruit for causal inference

Step 2
Emulate Target Trial protocol

- Eligibility criteria
- Treatment strategies
- Randomized assignment
- Start/End follow-up
- Outcomes
- Causal contrast
- Analysis plan

Hernán - 17th Armitage Lecture
Time zero of follow-up in the Target Trial

☐ For each person, 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

☐ Misalignment of eligibility criteria and treatment assignment leads to selection bias / immortal time bias
Misalignment of eligibility (E) and treatment assignment (A) prevents correct emulation

Hernán et al. *J Clin Epidemiol* 2016; 79:70-75
Why is it hard to align eligibility and treatment assignment at time zero?

- Time of eligibility may not be unique
  - An individual may meet the eligibility criteria at multiple times

- Treatment group may not be known at time zero
  - An individual’s treatment strategy/exposure plan will be revealed after time zero
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 Examples #1 (Hormone therapy) and #3 (Statins and cancer), eligibility criteria may be met at different times
  - What’s time zero?
EXAMPLE #4
Screening colonoscopy and colorectal cancer

- Colonoscopy screening recommended at age 50 and then every 10 years in the U.S.
  - 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
  - also, trials do not include older patients

- Need observational data to emulate a target trial
| **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 |
The observational data: U.S. Medicare

- Federal health insurance program for people 65 years or older, with disabilities or with ESRD
  - About 50 million enrollees per year
- Random sample of Medicare claims dataset, 1999-2012
  - Outpatient and inpatient services
  - Doctor services
  - Drug prescriptions
  - Screening colonoscopy since July 2001
- Medicare enrollees can meet eligibility criteria at multiple times
  - Every day since they turn 70 until 74
Choosing Time Zero when individuals meet eligibility at multiple times

Two unbiased choices:

- Choose a **single eligible time**
  - e.g., the first eligible time or a random eligible time

- Choose **every eligible time**
  - i.e., emulate a new trial starting at each eligible time
  - What we did for postmenopausal hormone therapy

Let’s do both for colonoscopy screening
Choosing a single eligibility time as time zero
Garcia-Albeniz et al. *Eur J Epid* 2017

1. Colonoscopy group: individuals who meet the eligibility criteria and receive a colonoscopy
   - time zero is the time of the colonoscopy

2. No colonoscopy group: individuals who meet the eligibility criteria and did not receive a colonoscopy at first eligibility
   - time zero is, say, their first eligible time
Choosing all eligible times as time zero

☐ Emulate a new target trial each week of follow-up
  ■ Time zero is different in each trial

☐ Include in the emulation of each trial all individuals who are eligible at its corresponding time zero

☐ Combine all target trials for a more precise estimation
  ■ Need to take into account that some individuals will contribute to the emulation of several trials

☐ Use a robust variance
Target trial: sequential emulation

Week 1
70th birthday

If eligible

Screening

No Screening
Target trial: sequential emulation

Week 1
70th birthday

If eligible

Screening

No Screening

Week 2
(+ 1 week)

If eligible

Screening

No Screening
Target trial: sequential emulation

Week 1
70th birthday

If eligible
Screening
No Screening

Week 2
(+ 1 week)

If eligible
Screening
No Screening

Repeat until reaching week 260
Pool data of all 260 “trials”
Multiple eligibility times

Single eligibility time
Both approaches are valid choices of time zero

- Because they respect the basic principle of study design
  - Time zero is the time when eligibility is met and treatment strategies are assigned

- Consider two alternative observational analyses that do not respect this principle
  - and therefore do not emulate a target trial
Incorrect emulation #1
Redefine the “No colonoscopy” group

1. Colonoscopy group: individuals who meet the eligibility criteria and receive a colonoscopy
   - time zero is the time of the colonoscopy

2. No colonoscopy group: individuals who meet the eligibility criteria and did not receive a colonoscopy at first eligibility during the follow-up
   - time zero is their first eligible time
Incorrect emulation #2
Select arbitrary time zero and look back

1. Colonoscopy group: individuals who meet the eligibility criteria and received a colonoscopy in the five years before time zero
   - time zero is, say, January 2010

2. No colonoscopy group: individuals who meet the eligibility criteria and did not receive a colonoscopy in the five years before time zero
   - time zero is, say, January 2010
Correct emulation

Incorrect emulation #2

Original analysis

- No screening
- Colonoscopy

Treatment before $t_0$, eligibility at $t_0$
Why is an incorrect time zero so frequent in observational analyses?

- Because epidemiologists and statisticians lost their way some time in the 1960-1970s
- They started to organize their analyses around the concept of **person-time**
- A typical analysis
  - decomposes each person in a collection of person-years
  - classifies each person-year as exposed or unexposed
  - compares risk of exposed and unexposed person-years
But there is no time zero for person-years

- Time zero is start of follow-up for a person
  - Not for a person-year

- Once person-years take over the analysis, time zero disappears from the picture
  - Observational analyses deviate from the target trial
  - Absolute risks cannot be correctly estimated

- Conducting analyses that revolve around person-time is like shooting ourselves in the foot
2 key components of the emulation of the target trial

1. Specification of time zero
   - synchronized with determination of eligibility and assignment of treatment strategies

2. Randomized assignment
   - Emulation requires adjustment for confounding

☐ Lack of randomization typically blamed for the failings of observational analyses, but...
   - confounding cannot be addressed until time zero is correct

☐ Suppose time zero is correct, can we really emulate a randomized assignment using observational data?
EXAMPLE #5
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
**Target trial: Statin therapy and coronary heart disease**

**Protocol summary**

<table>
<thead>
<tr>
<th><strong>Eligibility criteria</strong></th>
<th>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.</th>
</tr>
</thead>
</table>
| **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 for emulation: The Health Improvement Network (THIN)

- Primary care electronic health records
  - 6.2 million individuals from 350 general practices in the UK (2009)
- For each individual
  - demographic and socioeconomic characteristics
  - symptoms, diagnoses, referrals, lab results
  - prescriptions and procedures
  - some lifestyle information
  - vital status and cause of death data
Target trial emulation

- Use observational data from THIN to emulate the components of the target trial.
- Eligible individuals classified into:
  - Strategy 1 if they initiated statin treatment
  - Strategy 2 if they did not initiate statin treatment
  - During the baseline month
- Baseline month January 2000
  - 3178 individuals met all the eligibility criteria
  - 18 were initiators
  - 1 initiator developed CHD
Sequence of target trials
Emulation

- Emulate a target trial starting each calendar month between January 2000 and November 2006
  - 83 target trials with a 1-month enrollment period
- For each trial
  - Follow-up starts at the trial-specific baseline and ends at diagnosis of CHD, death, lost to follow-up, or January 2007
  - Eligibility criteria applied at each baseline
Sequence of target trials

Emulation

- 74,806 individuals eligible for at least 1 trial
- On average each eligible individual participated in the emulation of 11 trials
  - many non-initiators in the January 2000 trial still met all eligibility criteria in February
  - All 18 initiators in January 2000 were ineligible in February 2000 because they received treatment during the washout period for the February 2000 trial, and so on
<table>
<thead>
<tr>
<th>Adjustment variables</th>
<th>Often available in claims</th>
<th>Not generally available in claims</th>
</tr>
</thead>
<tbody>
<tr>
<td>□ Age, sex, calendar year</td>
<td>□ LDL cholesterol</td>
<td></td>
</tr>
<tr>
<td>□ Diagnoses: diabetes, hypertension...</td>
<td>□ HDL cholesterol</td>
<td></td>
</tr>
<tr>
<td>□ Drug use: beta-blockers, aspirin, hormone therapy...</td>
<td>□ Systolic blood pressure</td>
<td></td>
</tr>
<tr>
<td>□ Health care utilization: doctor visits, referrals, hospitalizations...</td>
<td>□ Body mass index</td>
<td></td>
</tr>
<tr>
<td></td>
<td>□ Smoking status</td>
<td></td>
</tr>
<tr>
<td></td>
<td>□ Alcohol use</td>
<td></td>
</tr>
</tbody>
</table>
Emulation of the target trial: Intention-to-treat effect and analysis

- Compare CHD incidence in initiators vs. no-initiators at baseline of each emulated trial
  - Regardless of their subsequent treatment
- Adjust for potential confounders
- Pool data across emulated trials to obtain a more precise effect estimate
  - Robust variance because of within-subject correlation
Hazard ratio (95% CI) of CHD THIN trials 2000-2006

<table>
<thead>
<tr>
<th></th>
<th>Intention-to-treat analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unique cases</td>
<td>635</td>
</tr>
<tr>
<td>Unique persons</td>
<td>74,806</td>
</tr>
<tr>
<td>Cases</td>
<td>6,335</td>
</tr>
<tr>
<td>Person-trials</td>
<td>844,800</td>
</tr>
<tr>
<td>Adjusted for age and sex</td>
<td>1.29 (1.06, 1.56)</td>
</tr>
<tr>
<td>Adjusted for all covariates</td>
<td>0.89 (0.73, 1.09)</td>
</tr>
</tbody>
</table>
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!
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; 175(4): 250-262
So does that mean that lack of randomization is Ok?

☐ No

☐ Confounding due to lack of randomization always possible when using observational real world data

☐ Explicitly emulating the target trial only eliminates self-inflicted injuries
  ■ Selection bias, immortal time bias...
  ■ Confounding is not a self-inflicting injury
A common misinterpretation

☐ You are saying that observational studies are as good as randomized trials?
  “This is a cohort study that tries to turn itself into a clinical trial. This involves a series of assumptions and manoeuvres which lack credibility.”
  Anonymous JAMA reviewer, April 2014

☐ No, the point is **not** that observational studies can turn themselves into randomized experiments
  They can’t
The point is that we can do better

- by using observational data to explicitly emulate randomized trials

- The limitations of observational studies remain
  - confounding, mismeasurement...
- but we do not compound them with additional problems
  - selection bias, immortal time bias...
Failures in the emulation of randomized assignment

- Treatments that are proxies for prognostic factors that remain unmeasured
  - Example: Preventive interventions (e.g., screening colonoscopy) and mortality
    - Garcia-Albeniz et al. *Am J Epidemiol* 2019
  - Unmeasured confounding: biased effect estimate

- Treatments that are universally administered to individuals with certain prognostic factors
  - Example: antihypertensives vs no antihypertensives
    - Danaei et al. *J Clin Epidemiol* 2018
  - Intractable confounding: biased effect estimate
The Target Trial concept leads to a simple two-step 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
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
Emulation of a target trial is what we do when we cannot conduct the trial

- Reasonable people will always prefer a randomized trial, but often there is no alternative to observational studies
  - we better keep improving them
  - because people will keep using observational data to guide their decisions
- And we have identified some simple ways of improving observational analyses to learn what works and what harms
The Target Trial concept leads to a simple two-step 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
Not explicitly describing our causal goal is like shooting without a target.
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
Thank you

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

☐ Causal Inference book
  ■ Free online, google “causal inference book”
  ■ www.facebook.com/causalinference