Designing analyses of healthcare databases to emulate randomized trials

Miguel Hernán
DEPARTMENTS OF EPIDEMIOLOGY AND BIOSTATISTICS

HARVARD T.H. CHAN SCHOOL OF PUBLIC HEALTH

@_MiguelHernan
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 medical 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)
Step 1
Specify Target Trial protocol

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

Step 2
Emulate Target Trial protocol

- Eligibility criteria
- Treatment strategies
- Randomized assignment
- Start/End follow-up
- Outcomes
- Causal contrast
- 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
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 2018
<table>
<thead>
<tr>
<th>Eligibility criteria</th>
<th>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.</th>
</tr>
</thead>
</table>
| 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, 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
  - No emulation of target trial

### Mortality hazard ratio (95% CI)

<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>
Immortal time bias in some observational studies

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

<table>
<thead>
<tr>
<th></th>
<th>Mortality hazard ratio (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>These studies</td>
</tr>
<tr>
<td>Cancer-specific</td>
<td>0.35 (0.27, 0.44)</td>
</tr>
<tr>
<td>All-cause</td>
<td>0.39 (0.33, 0.45)</td>
</tr>
</tbody>
</table>
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
Step 1
Specify Target Trial protocol

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

Choosing 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 - Target trial
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
Emulation of time zero is not straightforward when there are multiple eligibility times

- In Example #1 (Statins in cancer patients), eligibility criteria are met as a single time
  - Cancer diagnosis
  - That’s time zero

- In the next example eligibility criteria may be met at different times
Examples of Target Trial emulation using different types of observational data

1. Statins and mortality in cancer patients
   - Cancer registry + Insurance claims

2. Screening colonoscopy and cancer
   - Insurance claims

3. Epoetin therapy and mortality in dialysis patients
EXAMPLE #2
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 (but I didn’t explained here) in the postmenopausal hormone therapy example

Let’s do both for colonoscopy screening
Assign individuals who receive a colonoscopy while meeting the eligibility criteria to the colonoscopy strategy
- time zero is the time of the colonoscopy

Assign individuals who did not receive a colonoscopy at first eligibility to the no colonoscopy group
- time zero is their first eligible time during the study period
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
Sequential emulation

Enrollee turns 70
Eligible?

Week 1

No screening ——>

Screening ———

Hernán - Target trial
Sequential emulation

Enrollee turns 70
Eligible?

Week 1
No screening
Screening

70 +1 week
Eligible?

Week 2
No screening
Screening

70 +2 week
Eligible?

Week 3
No screening
Screening

Target trial
Sequential emulation

Enrollee turns 70

Week 1
No screening

Week 2
Screening
No screening

Week 3
Screening
No screening

Week 4
Screening
No screening

Week 260
No screening

Pooling of all the person-months from the 260 “trials”
These two approaches to choose time zero are valid

- 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
1. 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 group have little opportunity to be diagnosed
  - similar to naïve per-protocol analyses in randomized trials
Correct emulation

Incorrect emulation #1

Original analysis

No screening

Colonoscopy

Treatment at $t_0$, eligibility after $t_0$
2. 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
Correct emulation

Incorrect emulation #2

Original analysis

No screening

Colonoscopy

Treatment before \( t_0 \), eligibility at \( t_0 \)

Years

CRC cumulative incidence

0.05

0.04

0.03

0.02

0.01

0.00

0.00

0.01

0.02

0.03

0.04

0.05

0.00

0.01

0.02

0.03

0.04

0.05

0.00

0.01

0.02

0.03

0.04

0.05

0.00
Basic principle of trial design
(and of observational analyses that emulate a target trial)

- Treatment assignment and the determination of eligibility occur simultaneously at time zero

- Observational analyses that violated this principle yielded implausible estimates
  - Good news: correct time zero determination is always possible
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
All of the above was an oversimplification

- We compared **point** interventions at baseline
  - Initiation of statin therapy
  - Screening colonoscopy

- Most clinical practice is about treatment strategies that are **sustained** over time
  - Initiation of therapy + continuation of therapy in the absence of contraindications + possible adjustment of dose + ...
Examples of Target Trial emulation using different types of observational data

1. Statins and mortality in cancer patients
   - Cancer registry + Insurance claims

2. Screening colonoscopy and cancer
   - Insurance claims

3. Epoetin therapy and mortality in dialysis patients
   - Insurance claims + supplementary data
EXAMPLE #3
Epoetin therapy and mortality

- **Question**
  - What is the effect of different doses of epoetin on the mortality of hemodialysis patients?

- **Data: US Renal Data System**
  - Medicare claims database
  - ~18,000 eligible elderly patients
The target trial

- Eligibility criteria
  - End-stage renal disease, 3 months on hemodialysis

- Strategies
  - Fixed weekly dose of intravenous epoetin
    - 15,000, 30,000, or 45,000 units

- Follow-up
  - From 3 months after hemodialysis onset until death, loss to follow-up or administrative end of the study (1 year)

- Outcome
  - All-cause mortality

...
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

- Treatment-confounder feedback
  - Need Robins’s g-methods
There is treatment-confounder feedback because the time-varying confounder is affected by previous treatment

special statistical methods (g-methods) are needed: inverse-probability weighting, g-formula, etc
Survival under 3 epoetin dosing regimes

Zhang et al. CJASN 2009; 21:638-644
But this is a silly target trial

☐ In clinical practice, patients do not receive a fixed weekly dose of epoetin
   ■ That would be clinical malpractice

☐ Rather, actual clinical strategies are dynamic
   ■ A patient’s weekly dose depends on her hemoglobin or hematocrit, which in turn depends on her prior weekly dose
More reasonable strategies for a target trial

1. Mid-Hematocrit strategy
   - epoetin to maintain Hct between 34.5% and 39.0%

2. Low-Hematocrit strategy
   - epoetin to maintain Hct between 30.0% and 34.5%.

- Under both strategies, epoetin dose is
  - increased by >10% if previous Hct below target
  - decreased by <10% times [previous Hct minus lower end of range] or increased by <10% times [upper end of range minus Hct] if Hct within target
  - decreased by >25% if Hct above target
More reasonable strategies imply more work for the analyst

- Need to specify a more detailed protocol for the target trial

- Need to specify how to emulate that protocol
  - Appropriate adjustment for time-varying confounders becomes critical
  - Zhang et al. *Medical Care* 2014
<table>
<thead>
<tr>
<th>Components</th>
<th>Hypothetical Open-labeled, Nonblinded Randomized Clinical Trial</th>
<th>Emulated Trial Using USRDS Observational Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aim</td>
<td>To study the risks and benefits of epoetin therapy to target hematocrit (Hct) 34.5%-39.0% vs. 30.0%-34.5%</td>
<td>Same</td>
</tr>
<tr>
<td>Study population</td>
<td>Elderly ESRD patients with both diabetes and cardiovascular disease who initiated hemodialysis in US outpatient facilities between January 1, 2006 and December 31, 2008</td>
<td>Same</td>
</tr>
<tr>
<td>Eligibility criteria</td>
<td>Inclusion criteria: ≥65y of age, initiated outpatient dialysis within 90d of enrolling in the Medicare ESRD Program in 2006–2008, evidence of both diabetes and cardiovascular disease before or at baseline (end of the third month of hemodialysis)</td>
<td>Inclusion criteria: same</td>
</tr>
<tr>
<td></td>
<td>Evidence of diabetes was ascertained as an underlying cause of ESRD, Medical Evidence Form (MEF) reporting of diabetes (on insulin, with oral medications, or without medications) or diabetic retinopathy, or a hospitalization with (primary or secondary) ICD-9 code 252.x during the 3 mo before baseline. Evidence of cardiovascular disease was ascertained as MEF reporting of congestive heart failure, atherosclerotic heart disease, cerebrovascular disease, or peripheral vascular disease; or a hospitalization with (primary or secondary reason) ICD-9 codes 428.0 (congestive heart failure), 414.01 (atherosclerotic heart disease), 410-418 (cerebrovascular disease), or 413.9 (peripheral vascular disease) during the 3 mo before the start of follow-up. Evidence of cancer was also obtained from the MEF file.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Exclusion criteria: history of cancer before ESRD, no epoetin therapy in the first 30d of dialysis, in a nondialysis facility (e.g., hospital) at baseline, kidney transplantation or peritoneal dialysis before baseline, use of carboptcit before baseline, and hematocrit &lt;24% at baseline</td>
<td>Exclusion criteria: same</td>
</tr>
<tr>
<td></td>
<td>In addition, patients were excluded if they had incomplete baseline covariates (see below)</td>
<td></td>
</tr>
</tbody>
</table>
Follow-up

**Start:** after completing 3 mo of hemodialysis therapy

**End:** 6 mo after baseline, death, or dropout/lack of follow-up, whichever happens first

Treatment assignment

Patients are randomly assigned to one of the following 2 dynamic treatment strategies:

1. **Mid Hct strategy:** intravenous epoetin-α to achieve and maintain hematocrit values of 34.5%–≤39.0% or
2. **Low Hct strategy:** intravenous epoetin-α to achieve and maintain hematocrit values of 30.0%–≤34.5%

Under both strategies, monthly epoetin dose is changed according to the following rules:

i. if previous hematocrit is below the target range, epoetin dose is increased by ≥10%;

ii. if previous hematocrit is within target range, epoetin dose is decreased by ≤10% times (hematocrit minus lower end of range) or increased by ≤10% times (upper end of range minus hematocrit);

iii. if previous hematocrit is above the target range, epoetin dose is decreased ≥25% or withheld

The epoetin dose is left to the discretion of the treating physician during the month after the patient undergoes hemodialysis at a facility not participating in the study (e.g., hospital, hospice, nursing home or home health services) and after epoetin dose was withheld. The administration of IV iron is left to the discretion of the treating physician. Darbepoeitin is not made available to patients.

**Start:** same.

**End:** same. Dropouts/lack of follow-up defined as the earlier of

- key data become unreliable, eg, epoetin dose ≥0 even though Hct level was not reported in the claim
- 30-d gap in outpatient dialysis or inpatient claims
- switch to darbepoeitin (to emulate a trial in which darbepoeitin was not available)

Patients are classified as following one, both, or neither of the Mid Hct/Low Hct strategies

If a patient's treatment data during the first month of follow-up are consistent with the rules on the left for

- one strategy: the patient is assigned to that strategy
- both strategies: we clone the patient's data and assign each clone to one of the 2 strategies

- neither strategy: the patient is ineligible for the study

If a patient has >1 dialysis claim during 1 mo, only the data in first dialysis claim is used

(Continued)
Endpoints

Primary: all-cause mortality
Secondary: a composite endpoint of mortality and a hospitalization for MI, stroke, or congestive heart failure

Statistical analysis

**Intention-to-treat analysis:** Cox model with indicator for strategy (Mid or Low Hct) and with inverse-probability (IP) weights to adjust for selection bias due to loss to follow-up. Weighted survival curves under each strategy.

**Per-protocol analysis:** Patients are artificially censored when they deviate from their assigned strategy. IP weights for artificial censoring are estimated as a function of history of epoetin dose, hematocrit, change in hematocrit values, iron treatment, hospitalization, and product terms between these variables. Patients are not artificially censored if (a) the patient is not on dialysis or (b) there are no hematocrit values. Censoring is also required if the patient is not on dialysis or if the patient is not on iron treatment.

Primary: same
Secondary: same. Cardiovascular events are identified through ICD-9-CM codes for primary reason for hospitalization on Medicare hospital claims using the following International Classification of Disease, Ninth Revision, Clinical Modification (ICD-9-CM) codes: MI codes 410.xx (except 410.x1); CHF: codes 402.x1, 425.xx, 428.xx, 518.4, and 398.91; and stroke: codes 430.xx, 431.xx, 432.xx, 433.xx, and 434.xx.

**Intention-to-treat analysis:** same, except that Cox model also includes baseline variables (age at ESRD onset, race, sex, US geographic region, dialysis chain membership, patient BMI, Charlson Index score, tobacco use, drug/alcohol dependence, chronic obstructive pulmonary disease, serious conditions including anemia, incontinence, ability to ambulate, and inability to transfer, predialysis hematocrit, predialysis epoetin use, average baseline hemoglobin, average baseline epoetin dose, and average baseline iron dose). Survival curves are standardized to baseline variables.

**Per-protocol analysis:** same

BMI indicates body mass index; ESRD, end-stage renal disease; Hct, hematocrit; MI, myocardial infarction.
Survival under these 2 dynamic strategies
These examples show that successful emulation of a Target Trial requires

- High-quality data on
  - treatment
  - outcome
  - confounders

- 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
A common misinterpretation

☐ You are saying that observational studies are as good as RCTs?
  ■ “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...
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.
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