Beyond Controlling for Confounding: Design Strategies to Avoid Selection Bias and Improve Efficiency in Observational Studies

A Case Study of Screening Colonoscopy

September 27, 2018
Our Team

Xabier Garcia de Albeniz Martinez
+34 93.3624732
xgarcia@rti.org

Bradley Layton
919.541.8885
jblayton@rti.org
Poll Questions

Poll question 1

Ice breaker: Which of the following study designs is best to evaluate the causal effect of a medical intervention?

- Cross-sectional study
- Case series
- Case-control study
- Prospective cohort study
- Randomized, controlled clinical trial
We All Trust RCTs… Why?

• Obvious reasons
  – No confusion (i.e., exchangeability)

• Not so obvious reasons
  – Exposure represented at all levels of potential confounders (i.e., positivity)
  – Therapeutic intervention is well defined (i.e., consistency)
  – … And, because of the alignment of eligibility, exposure assignment and the start of follow-up (we’ll soon see why is this important)

RCT = randomized clinical trial.
Poll Questions

We just used the C-word: “causal” effect

Poll question 2

Which of the following is true?

- In pharmacoepidemiology, we work to ensure that drugs are effective and safe for the population
- In pharmacoepidemiology, we want to know if a drug *causes* an undesired toxicity
- Causal inference from observational data can be questionable, but being explicit about the *causal goal* and the validity conditions help inform a scientific discussion
- All of the above
In Pharmacoepidemiology, We Try to Infer Causes

- These are good times to be an epidemiologist thanks to the wealth of data available (claims, electronic medical records, wearable devices, etc.)

- This helps facilitate precision and positivity… but the identifiability conditions have little to do with how large our database is
What Are We Going to Talk About Today?

• Much of the focus in observational studies is placed on adjusting for confounding
  – This is **necessary** because we do not have control over the existence of common causes

• There’s still room for bias when some basic design strategies are not followed
  – Be sure to align eligibility, exposure assignment and the start of follow-up

• Letting people contribute to your analysis whenever they are eligible by emulating a series of trials
Identifiability Conditions

To identify a causal effect, we need (i) data and (ii) assumptions external to the data

1. The conditional probability of receiving every value of treatment, though not decided by the investigators, depends only on the measured covariates (i.e., *conditional exchangeability*, no unmeasured confounding).

2. The values of treatment under comparison correspond to well-defined interventions that, in turn, correspond to the versions of treatment in the data (i.e., *consistency*).

3. The conditional probability of receiving every value of treatment is greater than zero (i.e., *positivity*).

## When Can the Identifiability Conditions Fail?

<table>
<thead>
<tr>
<th>Identifiability Condition</th>
<th>RCT</th>
<th>Observational Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exchangeability</td>
<td>Losses to follow-up do not happen at random</td>
<td>We miss baseline confounders</td>
</tr>
<tr>
<td></td>
<td>Losses to follow-up do not happen at random</td>
<td></td>
</tr>
<tr>
<td>Positivity</td>
<td>Artificial assignment of treatment guarantees it</td>
<td>Data are sparse or there are too many strata</td>
</tr>
<tr>
<td>Consistency</td>
<td>Protocol does not specify accurately the experimental intervention, or researchers do not follow it</td>
<td>Intervention is not well-defined or database does not differentiate multiple versions of the exposure (e.g., prevalent users)</td>
</tr>
</tbody>
</table>
Why We (Think We) Love RCTs

Confounding

L → A → Y

Age → PPI → Death

• Much of the attention in observational research is focused on managing confounding, which is fine

• How to handle confounding
  – Measure the common causes and use your favorite adjustment method
  – Randomize

PPI = proton pump inhibitor.
Let’s Assume We Have Confounding and Positivity Under Control… What Else Can Go Wrong?

- Problems derived from the lack of synchronization in time of eligibility, treatment assignment, and time zero:
  - Time of eligibility (E): point in time when patients meet the eligibility criteria
  - Treatment assignment (A): point in time when patients are classified into exposure groups
  - Time zero (T₀): point in time when follow-up starts

- This is not a problem in RCT because of the following:
  - Time of eligibility: when deemed eligible by the PI
  - Treatment assignment: randomization happens shortly afterwards
  - Time zero: date of randomization.
Is This a Problem in Observational Studies?

Some “EPI-101” biases are a consequence of this lack of synchronization

- **Classical immortal time bias**
  - Information on treatment after time zero is used to assign individuals to a treatment strategy. This time is immortal time.
  - The definition of A by looking into the future guarantees that individuals are alive for that period of time
Is This a Problem in Observational Studies?

Some “EPI-101” biases are a consequence of this lack of synchronization

- **Prevalent/current user bias**
  - Inclusion of individuals who initiated the exposure of interest some time before the start of follow-up
  - Prevalent users have survived the drug for a period of time
  - Past use of the drug can affect baseline covariates
  - Cannot inform health policy (i.e., cannot prescribe to be a “prevalent user”)

![Diagram](image-url)
Let’s Continue the Conversation
With a Case Study…
CRC Screening: Intro

• CRC screening can prevent cancer

• CRC screening tools
  – Fecal occult blood test
  – Sigmoidoscopy
  – Colonoscopy

• RCTs have proved the following:
  – Sigmoidoscopy (either as a single intervention or twice in 3-5 years) reduces CRC incidence and CRC mortality

• No RCTs for colonoscopy (yet)

CRC = colorectal cancer.
Research Question

• What is the effectiveness of screening colonoscopy in individuals aged 70-74?
  – Population barely (if at all) represented in ongoing RCTs
  – Over a decade of screening colonoscopy use in Medicare

• Let’s see
  – What challenges there are to answering this question using observational data (administrative data set)
  – How to deal with those challenges
Subtleties Specific to This Research Question

That we learned from sigmoidoscopy trials

• Screening sigmoidoscopy has very **little/no effect** on all-cause mortality:
  – RR = 0.98 (95% CI, 0.96-0.99)\(^1\)

• The effect of screening sigmoidoscopy in CRC incidence is nonmonotonic.

Subtleties Specific to This Research Question

That we learned from sigmoidoscopy trials

- **All-cause mortality** is prone to be *confounded* more than CRC incidence in an observational setting
Once-Only Sigmoidoscopy in Colorectal Cancer Screening: Follow-up Findings of the Italian Randomized Controlled Trial—SCORE

Nereo Segrè, Paola Armanoli, Luigina Bonelli, Mauro Risio, Stefania Scialiero, Marco Zappa, Bruno Andreoni, Arrigo Arrigoni, Luigi Bisenti, Claudia Castello, Cristiano Croste, Fabio Falcini, Franco Ferrero, Adriano Glaconin, Orietta Giuliani, Alessandra Santarelli, Carmen Beatriz Visoli, Roberto Zenetti, Wendy S. Atkin, Carlo Senore; and the SCORE Working Group

Lancet 375:1624

Once-only flexible sigmoidoscopy screening in prevention of colorectal cancer: a multicentre randomised controlled trial

Wendy S Atkin, Rob Edwards, Ines Kroji-Hans, Kate Wooldrage, Andrew R Hart, John M A Northover, D Max Parkin, Jane Wardle, Stephen W Duffy, Jack Curick, UK Flexible Sigmoidoscopy Trial Investigators

Control and intervention groups

Control, screened, and not screened groups
Effect of Flexible Sigmoidoscopy Screening on Colorectal Cancer Incidence and Mortality: A Randomized Clinical Trial

All-cause mortality

HR: 0.97 (95% CI, 0.93-1.02)

HR = hazard ratio.
Subtleties Specific to This Research Question

Decisions

- **All-cause mortality** is more prone to be **confounded** than CRC incidence in an observational setting
  - Stick to CRC incidence and stage at diagnosis
  - These are relevant clinical outcomes and plausibly the main mediators in improving cancer-specific survival

- The effect of screening colonoscopy in **CRC incidence is nonmonotonic**.
  - Use a cohort design to plot **cumulative incidence curves**
  - Standardize the cumulative incidence curves using a discrete-time hazards model¹
  - Estimate the **absolute difference** at the end of the follow-up

First Step: Emulation of the Target Trial

<table>
<thead>
<tr>
<th>Component</th>
<th>Target Trial</th>
<th>Emulated Trial Using Medicare</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Aim</strong></td>
<td>To estimate the effect of screening colonoscopy on the 8-year risk of CRC</td>
<td>Same</td>
</tr>
<tr>
<td><strong>Eligibility</strong></td>
<td>Individuals aged 70-74&lt;br&gt;No previous CRC, adenomas, IBD, or screening in previous 5 years</td>
<td>Same, plus continuous enrollment in Medicare</td>
</tr>
</tbody>
</table>
| **Treatment strategies** | 1. Screening colonoscopy at baseline  
2. No screening for CRC at baseline | Same                                                               |
| **Outcome**        | CRC diagnosis within 8 years                                                 | Same                                                               |
| **Causal contrast**| Intention-to-treat effect  
Per-protocol effect                                                                 | Observational analog of a per-protocol effect                      |

IBD = inflammatory bowel disease.
Let’s Assume We Have Confounding Under Control: What Else Can Go Wrong?

**Design No. 1**

- We choose a **calendar date** (e.g., January 1, 2004) as an arbitrary time zero to start the follow-up and apply eligibility criteria
  - We assign eligible individuals to the colonoscopy arm if they received a colonoscopy in the previous 5 years or to the no colonoscopy arm otherwise

- Colonoscopies performed before time zero can affect eligibility (e.g., by detecting adenomas, tumors, IBD). This is an example of selection bias.
Let’s Assume We Have Confounding Under Control: What Else Can Go Wrong?

Design No. 1

- What do we find when we run this analysis in Medicare (1999-2012)?
  - Screening looks implausibly beneficial during the whole follow-up (we were expecting the detection of asymptomatic tumors at baseline)
Let’s Assume We Have Confounding Under Control: 
What Else Can Go Wrong?

**Design No. 2**

- Instead of choosing a calendar date as the anchor date:
  - Colonoscopy arm: eligible individuals who receive a colonoscopy, $t_0 = \text{time of colonoscopy}$
  - No screening arm: eligible individuals who do not receive a colonoscopy during the whole follow-up, $t_0$ being their first eligible time

- Most of the CRCs are diagnosed with a colonoscopy, thus individuals in the no screening group do not have an opportunity to have a CRC diagnosed. This is another example of selection bias.
Let’s Assume We Have Confounding Under Control: *What Else Can Go Wrong?*

**Design No. 2**

- No screening looks implausibly beneficial during the whole follow-up
Let’s Assume We Have Confounding Under Control: *What Else Can Go Wrong?*

Problems with the synchronization of the application of eligibility criteria, exposure assignment and the start of follow-up

**Design No. 3**

- Colonoscopy arm: eligible individuals who receive a colonoscopy, $t_0 = \text{time of colonoscopy}$.

- No screening arm: individuals who do not receive a colonoscopy at first eligibility time, $t_0 = \text{first eligible time}$.

- This approach appropriately emulates the target trial, no selection bias

- If first eligibility happens earlier than the first colonoscopy, this can unbalance groups (e.g., younger individuals in the control group).
Let’s Assume We Have Confounding Under Control: What Else Can Go Wrong?

Problems with the synchronization of the application of eligibility criteria, exposure assignment and the start of follow-up

**Design No. 3**

- Graph more similar to sigmoidoscopy RCTs
Let’s Assume We Have Confounding Under Control: What Else Can Go Wrong?

Problems with the synchronization of the application of eligibility criteria, exposure assignment and the start of follow-up

To summarize:

<table>
<thead>
<tr>
<th>Design</th>
<th>Treatment Assigned</th>
<th>Eligibility Determined</th>
<th>Individuals Used More Than Once</th>
<th>Arm</th>
<th>N</th>
<th>CRC Cases</th>
<th>8 Year Risk Difference, % (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Before $t_0$</td>
<td>At $t_0$</td>
<td>No</td>
<td>No screening</td>
<td>6,507</td>
<td>178</td>
<td>Ref.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Screening</td>
<td>37,844</td>
<td>492</td>
<td>$-1.7$ (–2.2, –1.3)</td>
</tr>
<tr>
<td>2</td>
<td>At $t_0$</td>
<td>After $t_0$</td>
<td>No</td>
<td>No screening</td>
<td>6,241</td>
<td>11</td>
<td>Ref</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Screening</td>
<td>46,872</td>
<td>685</td>
<td>$1.7$ (1.4, 2.1)</td>
</tr>
<tr>
<td>3</td>
<td>At $t_0$</td>
<td>At $t_0$</td>
<td>No</td>
<td>No screening</td>
<td>72,249</td>
<td>1,086</td>
<td>Ref</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Screening</td>
<td>46,872</td>
<td>685</td>
<td>$-0.67$ (–1.03, –0.28)</td>
</tr>
</tbody>
</table>
Let’s Assume We Have Confounding Under Control: 
*What Else Can Go Wrong?*

Problems with the synchronization of the application of eligibility criteria, exposure assignment and the start of follow-up

**Design No. 4**

The closest we can get to the emulation of a clinical trial

- We choose an anchor date (e.g., January 22, 2004) for eligibility and \( t_0 \); we synchronize exposure assignment with that date (this would be the equivalent of the randomization date)
- Colonoscopy arm: individuals who receive a **colonoscopy** in the next 7 days of the anchor date
- No screening arm: individuals who do **not** receive a colonoscopy in the next 7 days (increasing the number of days can take us back to situation No. 2)

- No selection bias
- In our database, only 56 eligible individuals had a screening colonoscopy at that index date, with only 2 CRC diagnoses during the follow-up
- These small numbers preclude the standardization of survival curves or obtaining precise estimates
Can We Make It Better?

Emulation of a sequence of target trials

- One way to increase efficiency is to emulate a sequence of target trials, starting at every interval during the study period. (i.e., same as situation No. 4 but at every available time interval)

- Let’s say that this is our Medicare population, and we decide to implement an RCT at time interval 3

Legend:

- Eligible person-time
- Ineligible person-time because of no Medicare enrollment.
- Ineligible person-time because of other exclusion criteria
- Eligible person-time, colonoscopy
- Eligible person-time, outcome
Can We Make It Better?

This is what we just did in situation No. 4

- We decide to implement an RCT at time interval 3

Legend:
- Eligible person-time
- Ineligible person-time because of no Medicare enrollment.
- Ineligible person-time because of other exclusion criteria
- Eligible person-time, colonoscopy
- Eligible person-time, outcome
Can We Make It Better?

We can continue emulating trials over time to increase efficiency

- We decide to run an RCT at time interval 3… and then another one at time interval 4

**Legend:**

- Eligible person-time
- Ineligible person-time because of no Medicare enrollment.
- Ineligible person-time because of other exclusion criteria
- Eligible person-time, colonoscopy
- Eligible person-time, outcome
We can continue emulating trials over time to increase efficiency

- We decide to implement an RCT at time interval 3… and then another one at time interval 4… and at all time intervals

**Can We Make It Better?**

**Subtleties:**
- A single person can contribute to several trials (e.g., id 05 contributes to trials 1-4)
- A single person can contribute to both arms (e.g., id 05 contributes to the “no colonoscopy” arm in trials 1-3 and to the “colonoscopy” arm in trial 4).
- Baseline characteristics are extracted at each baseline (e.g., id 06 has a baseline value of 1 for trials 0-3 and a baseline value of 2 for trials 4-6)
Can We Make It Better?

We can continue emulating trials over time to increase efficiency

- We decide to run an RCT at time interval 3… and then another one at time interval 4

ID 05 contributes to 2 cohorts and both arms

ID 06 contributes to 2 cohorts as a comparator
What Is Gained With This Additional Complexity?

Precision

- In this specific scenario of Medicare claims (screening colonoscopy and CRC incidence), allowing the individuals to contribute to multiple emulated trials was the equivalent of increasing the sample size of the unexposed group approximately tenfold

<table>
<thead>
<tr>
<th>Design</th>
<th>Treatment Assigned</th>
<th>Eligibility Determined</th>
<th>Individuals Used More Than Once</th>
<th>Arm</th>
<th>N</th>
<th>CRC Cases</th>
<th>Difference, % (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Before ( t_0 )</td>
<td>At ( t_0 )</td>
<td>No</td>
<td>No screening</td>
<td>6,507</td>
<td>178</td>
<td>Ref</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Screening</td>
<td>37,844</td>
<td>492</td>
<td>(-1.7) (-2.2, -1.3)</td>
</tr>
<tr>
<td>2</td>
<td>( t_0 )</td>
<td>After ( t_0 )</td>
<td>No</td>
<td>No screening</td>
<td>6,241</td>
<td>11</td>
<td>Ref</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Screening</td>
<td>46,872</td>
<td>685</td>
<td>1.7 (1.4, 2.1)</td>
</tr>
<tr>
<td>3</td>
<td>( t_0 )</td>
<td>( t_0 )</td>
<td>No</td>
<td>No screening</td>
<td>72,249</td>
<td>1,086</td>
<td>Ref</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Screening</td>
<td>46,872</td>
<td>685</td>
<td>(-0.67) (-1.03, -0.28)</td>
</tr>
<tr>
<td>4</td>
<td>( t_0 )</td>
<td>( t_0 )</td>
<td>Yes</td>
<td>No screening</td>
<td>1,762,816</td>
<td>21,954</td>
<td>Ref</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Screening</td>
<td>46,872</td>
<td>685</td>
<td>(-0.63) (-0.83, -0.43)</td>
</tr>
</tbody>
</table>
What Is Gained With This Additional Complexity?

Precision

• In this specific scenario of Medicare claims (screening colonoscopy and CRC incidence), allowing the individuals to contribute to multiple emulated trials was the equivalent of increasing the sample size of the unexposed group approximately tenfold.
Take-Home Messages

• Confounding happens, we do not have control over it
  – Much of the focus in observational studies is placed on adjusting for confounding

• Selection bias can be a self-inflicted injury
  – Align eligibility, exposure assignment and the start of follow-up

• Let units of observation contribute to your analysis whenever they are eligible by emulating a series of trials to increase efficiency
Further Reading

Methods


• Hernan MA, Robins JM. Using Big data to emulate a target trial when a randomized trial is not available. Am J Epidemiol. 2016;183(8):758-64.


Applications


Thank You
Questions?
Generating knowledge and providing greater understanding so that you—and those who regulate, pay for, prescribe, and use your products—can make better decisions.
Contact Us

Xabier Garcia de Albeniz Martinez
+34 93.3624732
xgarcia@rti.org

Bradley Layton
919.541.8885
jblayton@rti.org