External controls:
Emulating half of the target trial

Miguel Hernán

CAUSALab
A Center to Learn What Works

HARVARD
T.H. CHAN
SCHOOL OF PUBLIC HEALTH

Other relevant engagements

Adigens Health
Co-founder,
Senior Scientific Advisor

ADIA Lab
Advisory Board Member

Annals of Internal Medicine
Methods Editor

Karolinska Institutet
Principal Researcher

ProPublica
Adviser

Causal inference is the foundation of decision making

- Regulatory decisions require evidence of the comparative effectiveness and safety of medical interventions

- How do we make those causal inferences?

1. Randomized trials
   - Data generated for causal inference
2. Observational emulations of randomized trials
   - Data repurposed (or, less often, generated) for causal inference

But randomized trials can't possibly answer all questions

expensive unethical impractical untimely




- or too many populations of interest, or too many outcomes
- or simply nobody is willing to fund it (head-to-head trials?)

- The question isn't WHETHER to use observational data, but HOW to use observational data in the best possible way

Types of observational data

Research data

- Generated specifically for research
  - Cohort, case-control, case-crossover studies…
  - Biobanks
  - Disease registries
  - Randomized trials
  - …

Found data

- Generated for non-research purposes
- Repurposed for research
  - Electronic health records
  - Insurance claims databases
  - National registers
  - …

"Real world data"

"Routinely collected data"

The Target Trial 

- The (hypothetical) randomized trial that we would like to conduct to answer a causal question
- 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 question is not whether we're emulating a target trial, but whether we make our target trial explicit

The Target Trial 

- **Suggested more or less explicitly by many authors**
  - Dorn (1953), Wold (1954), 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 the causal question** (point at the Target)
   - Specify the protocol of the Target Trial
     - specify what we want to estimate = the estimand

2. **Answer the causal question** (shoot the Target)
   - Option A: Conduct the Target Trial
   - Option B : Use observational data to explicitly emulate Target Trial
     - Then apply appropriate causal inference analytics

Step 1
Specify Target Trial protocol

- Eligibility criteria
- Treatment strategies
- Assignment
- Outcomes
- Start/end of follow-up
- Causal contrasts

Causal estimand

- Identifying assumptions
- Data analysis

The ICH E9 Estimand framework uses different terminology for the same estimands

Target trial protocol

- Eligibility criteria
- Treatment strategies
- Assignment
- Outcomes
- Start/end of follow-up
- Causal contrasts
- Identifying assumptions
- Data analysis

- **Causal contrats**
  - Intention-to-treat effect (treatment policy effect)
  - Per-protocol effect (under other names)
- **"Intercurrent events"**
  - combine elements of the treatment strategies with competing events (causal contrast)

Step 1 Specify Target Trial protocol	Step 2 Emulate Target Trial protocol
Eligibility criteria	Data mapping for eligibility criteria
Treatment strategies	Data mapping for each component
Assignment	Data mapping for assignment
Outcomes	Data mapping for outcomes
Start/end of follow-up	Same
Causal contrasts	Observational analogs of contrasts
Identifying assumptions	Identifying assumptions
Data analysis	Data analysis

The target trial has to be emulatable

Eligibility criteria	Data mapping for eligibility criteria
Treatment strategies	Data mapping for each component
Assignment	Data mapping for assignment
Outcomes	Data mapping for outcomes
Start/end of follow-up	Same
Causal contrasts	Observational analogs of contrasts
Identifying assumptions	Identifying assumptions
Data analysis	Data analysis

External control arms are used to emulate one of the treatment strategies

Target trial protocol

- Eligibility criteria
- Treatment strategies
- Assignment
- Outcomes
- Start/end of follow-up
- Causal contrasts
- Identifying assumptions
- Data analysis

Data under one treatment strategy is generated by a non-randomized experiment

Data for the "control" treatment strategy is repurposed from an existing observational source

Emulations of target trials that combine experimental and observational (real world) data

- Experimental data
  - Generation is under the investigators' control
- Observational data
  - Generation is not under the investigators' control
  - Data must be repurposed to match what the investigators decided

- Challenge
  - the features of the experimental arm may not be emulatable
- Opportunity
  - The experimental arm may be designed to be emulatable

Each component of the target trial has to be emulatable

Eligibility criteria	Data mapping for eligibility criteria
Treatment strategies	Data mapping for each component
Assignment	Data mapping for assignment
Outcomes	Data mapping for outcomes
Start/end of follow-up	Same
Causal contrasts	Observational analogs of contrasts
Identifying assumptions	Identifying assumptions
Data analysis	Data analysis

Sufficient information on confounders is needed in both arms

Eligibility criteria	Data mapping for eligibility criteria
Treatment strategies	Data mapping for each component
Assignment	Data mapping for assignment
Outcomes	Data mapping for outcomes
Start/end of follow-up	Same
Causal contrasts	Observational analogs of contrasts
Identifying assumptions	Identifying assumptions
Data analysis	Data analysis

Emulations of target trials with external controls
In practice, 2 common situations:

- **External controls are included in the original design**
  - Preferred (should be encouraged)
  - Investigators can design an emulatable experimental arm
    - e.g., only eligibility criteria and outcomes that can be emulated with RWD, intensity of monitoring similar to that in the real world…
  - and with sufficient data on confounders
- **External controls are added after the fact**
  - Less preferred (should be discouraged)
  - Problematic if components of experimental arm aren't emulatable or insufficient data on confounders

External control added after the fact may work in special settings

- Very high adherence to treatment
- Outcomes unaffected by monitoring
- Few losses to follow-up
- Rich clinical data on treatment indications
- ...

JAMA Network Open

JAMA Network Open. 2025;8(3):e252152

Original Investigation | Oncology

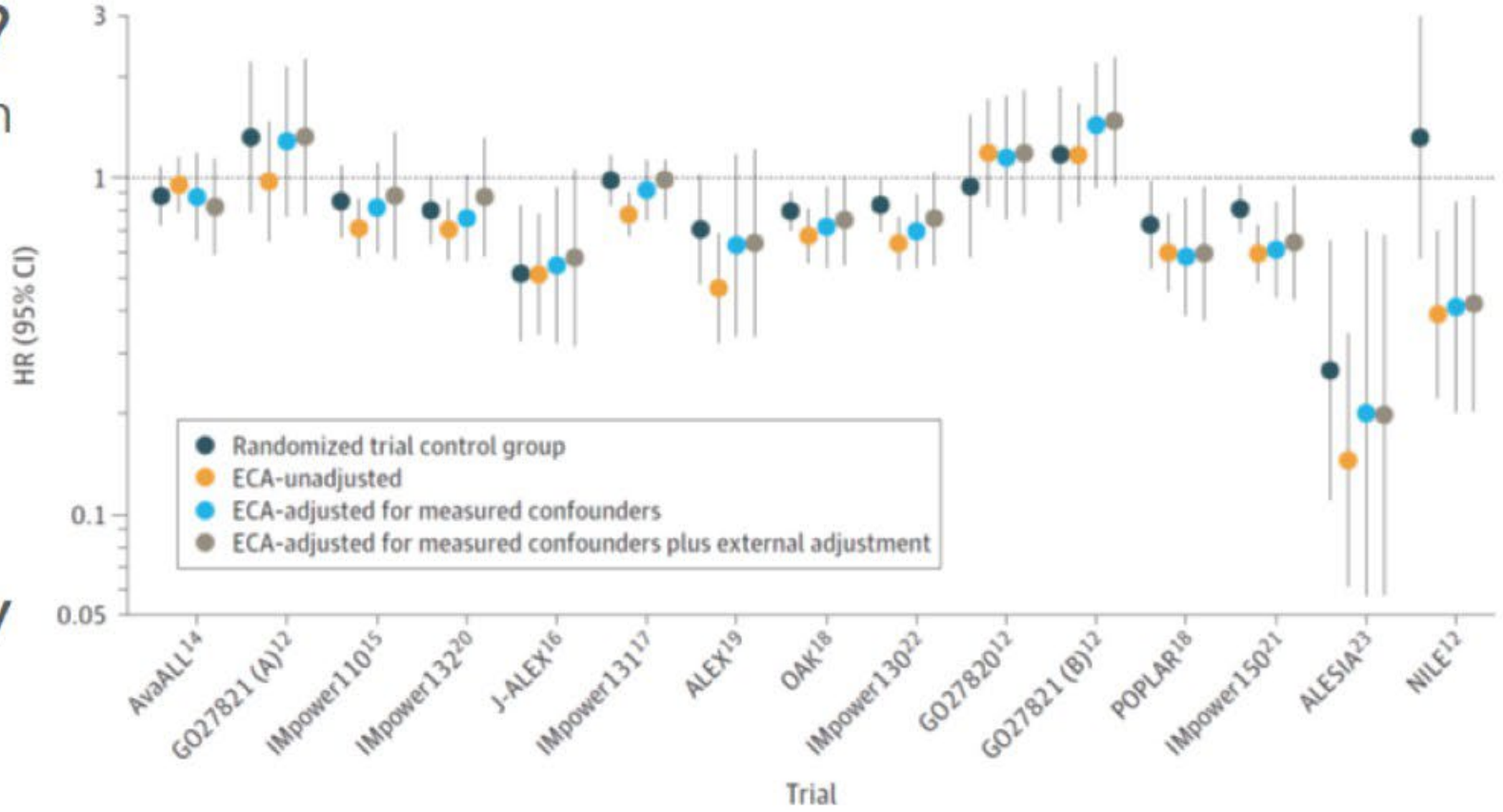
Quantitative Bias Analysis for Single-Arm Trials With External Control Arms

Alind Gupta, PhD; Grace Hsu, MSc; Seamus Kent, PhD; Stephen J. Duffield, PhD; Evie Merinopoulou, MSc; Alexandre Lockhart, PhD; Paul Arora, PhD; Joshua Ray, MSc; Samantha Wilkinson, PhD; Nicolas Scheuer, MSc; Sreeram V. Ramagopalan, PhD; Rolf H. H. Groenwold, PhD; Sanjay Popat, MBBS; Miguel A. Hernán, MD, DrPH

Example: Treatment of advanced non-small cell lung cancer

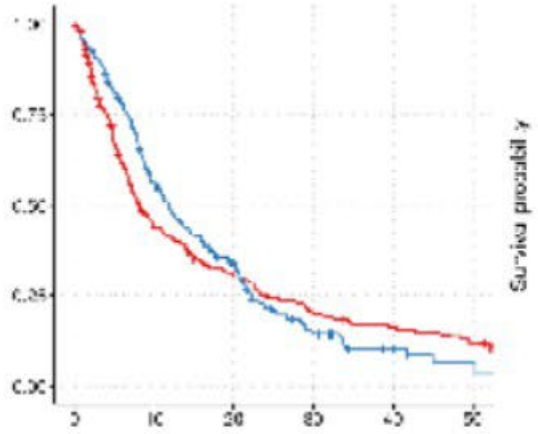
- How do we know it worked?
  - Similar results in 14 of 15 randomized trials after replacing placebo control by external control
- In practice, how do we identify the 15th trial?

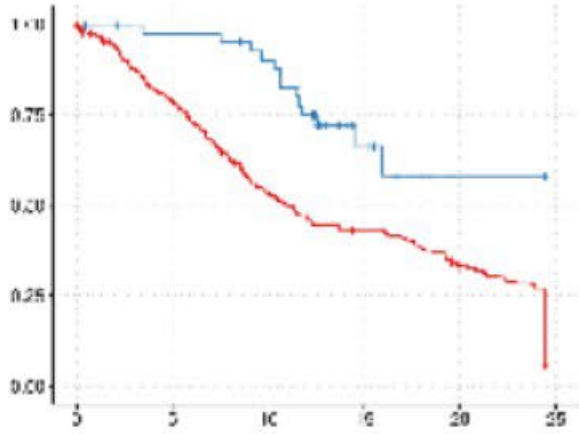
Figure 1. Hazard Ratio (HR) Estimates and 95% CIs



Diagnosing emulation failure when using external controls?

- **Benchmarking with internal control arm**
  - If an internal control arm exists (hybrid designs)
    - The internal arm may have short duration of follow-up
  - Benchmark external control with internal control arms
    - then extend the inference to longer follow-up
- **Negative controls**
  - e.g., instantaneous treatment effects?
- ...





Why external controls? Justifications need to be explicit (and probably regulated)

- **Feasibility**
  - rare disease, rapidly evolving therapeutical options, ethical constraints
- **Precision**
  - hybrid designs with both internal and external controls
- ~~**Convenience**~~
  - not a valid justification given the dangers of observational emulations

- **Normalizing the use of external controls cannot happen at the expense of fewer randomized trials**
  - Unacceptable to use external controls when internal controls would be appropriate

Target trial emulation is not a method, but a unifying framework

to increase transparency in

- **The articulation of the causal question**
  - Specification of the protocol of the target trial
- **The procedures followed to answer it**
  - Emulation of the target trial (mapping from protocol to data)
- **The reporting of the results**
  - TARGET guidelines (*JAMA*, *BMJ* 2025)
- **The evaluation of possible biases**
  - ROBINS-I: Cochrane Collaboration's tool for Risk of Bias (*BMJ* 2016)

Our research is funded by

- National Institute of Allergy and Infectious Diseases
- National Institute of Mental Health
- National Institute of Child Health and Development
- National Cancer Institute
- PCORI – Patient-Centered Outcomes Research Institute
- U.S. Department of Veterans Affairs
- U.S. Department of Defense
- Novo Nordisk

miguelhernan.org

CAUSALab
A Center to Learn What Works

HARVARD
T.H. CHAN
SCHOOL OF PUBLIC HEALTH

causalab.org

@CAUSALab