Steps for Emulating Target Trials

Bianca L De Stavola

Great Ormond Street Institute of Child Health, University College London

b.destavola@ucl.ac.uk

Work funded by MRC Methodology Grant: MR/R025215/1

One-day Course on Target Trial Emulation

28th November 2022

Overview

Limitations of RCTs

We have learnt from the previous lecture that conducting a RCT to address certain questions:

- Might not be ethical
- Might be impractical/expensive
- May be (practically) possible on selected populations
- Might not be timely

Limitations of RCTs

We have learnt from the previous lecture that conducting a RCT to address certain questions:

- Might not be ethical
- Might be impractical/expensive
- May be (practically) possible on selected populations
- Might not be timely

Limitations of RCTs

We have learnt from the previous lecture that conducting a RCT to address certain questions:

- Might not be ethical
- Might be impractical/expensive
- May be (practically) possible on selected populations
- Might not be timely

Limitations of RCTs

We have learnt from the previous lecture that conducting a RCT to address certain questions:

- Might not be ethical
- Might be impractical/expensive
- May be (practically) possible on selected populations
- Might not be timely

Observational studies

In contrast, observational studies allow addressing questions concerning observed treatments/interventions as received under real world conditions.
However,

- Addressing causal questions using observational data requires dealing with multiple possible sources of bias:
  - Selection (into the study) bias
  - Confounding bias (aka âselection into-treatmentâ bias)
  - Measurement error/missing data bias
- Additionally, other challenges/errors may arise from incorrect manipulation of the data, *e.g.* (see lecture 1):
  - time 0 bias
  - definition of the exposure groups
- Ideally, we should use the same design steps, whether the data are experimental or observational

Observational studies

In contrast, observational studies allow addressing questions concerning observed treatments/interventions as received under real world conditions.
However,

- Addressing causal questions using observational data requires dealing with multiple possible sources of bias:
  - Selection (into the study) bias
  - Confounding bias (aka âselection into-treatmentâ bias)
  - Measurement error/missing data bias
- Additionally, other challenges/errors may arise from incorrect manipulation of the data, *e.g.* (see lecture 1):
  - time 0 bias
  - definition of the exposure groups
- Ideally, we should use the same design steps, whether the data are experimental or observational

Observational studies

In contrast, observational studies allow addressing questions concerning observed treatments/interventions as received under real world conditions.
However,

- Addressing causal questions using observational data requires dealing with multiple possible sources of bias:
  - Selection (into the study) bias
  - Confounding bias (aka âselection into-treatmentâ bias)
  - Measurement error/missing data bias
- Additionally, other challenges/errors may arise from incorrect manipulation of the data, *e.g.* (see lecture 1):
  - time 0 bias
  - definition of the exposure groups
- Ideally, we should use the same design steps, whether the data are experimental or observational

Target Trial Emulation

- Hernàn and Robins (2016): propose a formal approach to adopting the same design principles of RCTs in research based on observational data

American Journal of Epidemiology

Vol. 183, No. 8
DOI: 10.1093/aje/kwv254
Advance Access publication:
March 18, 2016

Practice of Epidemiology

Using Big Data to Emulate a Target Trial When a Randomized Trial Is Not Available

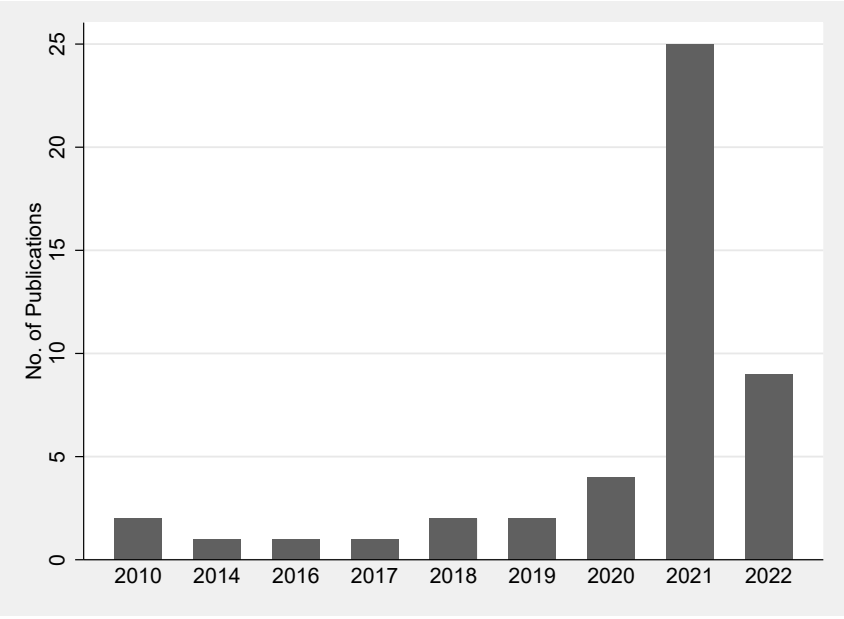
Miguel A. Hernán* and James M. Robins

* Correspondence to Dr. Miguel A. Hernán, Department of Epidemiology, 677 Huntington Avenue, Boston, MA 02115 (e-mail: miguel_hernan@post.harvard.edu).

Initially submitted December 9, 2014; accepted for publication September 8, 2015.

Target Trial Emulation

- Popularity of TTE increasing (Web of Science May 2022):



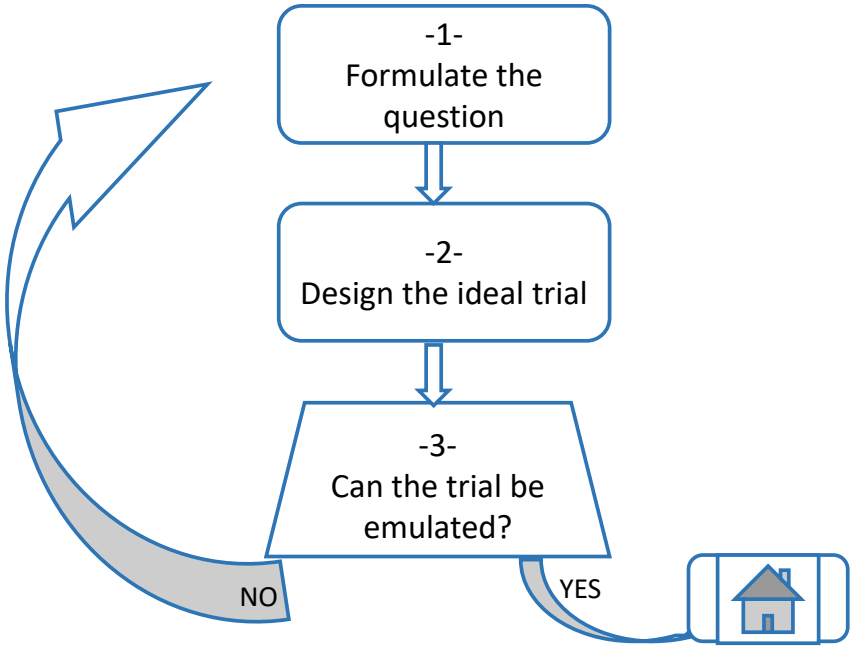
Target Trial Emulation

- Popularity of TTE increasing (Web of Science May 2022):

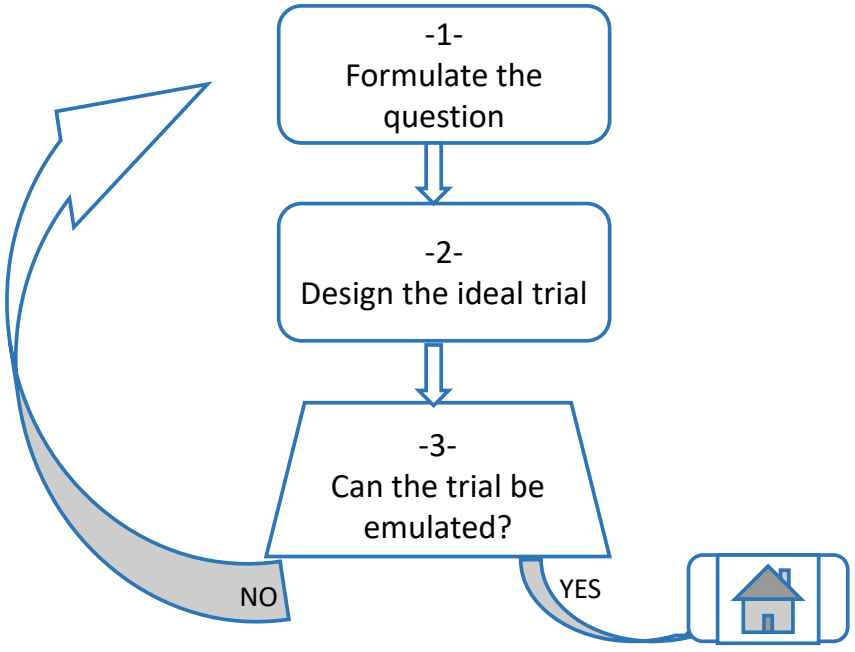
Most recent publications in 2022:

Becker et al	Cannabis use, pain interference, and prescription opioid receipt among persons with HIV: a target trial emulation study
Madenci et al	Strengthening Health Services Research Using Target Trial Emulation: An Application to Volume-Outcomes Studies
Bakker et al	Analysing electronic health records: The benefits of target trial emulation
Admon et al	Emulating a Novel Clinical Trial Using Existing Observational Data Predicting Results of the PreVent Study
Garcia-Albeniz et al	The value of explicitly emulating a target trial when using real world evidence: an application to colorectal cancer screening
Matthews et al	Comparing Effect Estimates in Randomized Trials and Observational Studies From the Same Population: An Application to Percutaneous Coronary Intervention
Trevisan et al	Stopping mineralocorticoid receptor antagonists after hyperkalaemia: trial emulation in data from routine care
Rossides et al	Infection risk in sarcoidosis patients treated with methotrexate compared to azathioprine: A retrospective 'target trial' emulated with Swedish real-world data
Hernan et al	Specifying a target trial prevents immortal time bias and other self-inflicted injuries in observational analyses
Lyu et al	Arteriovenous Access Type and Risk of Mortality, Hospitalization, and Sepsis Among Elderly Hemodialysis Patients: A Target Trial Emulation Approach
Reitblat et al	Radical prostatectomy versus external beam radiation therapy for high-grade, clinically localized prostate cancer: Emulation of a target clinical trial
Wu et al	Validation of Machine Learning-Based Individualized Treatment for Depressive Disorder Using Target Trial Emulation

Target Trial Emulation in Practice



Target Trial Emulation in Practice



Because of the observational nature of the data, the target trial can only be a pragmatic trial*

* A trial of an intervention in the general population delivered by general practitioners and unblindly

Step 2: the Ideal Target Trial

A protocol for the ideal target trial is specified:

1. *Eligibility criteria*	. . . . .
2. *Recruitment period*	. . . . .
3. *Follow-up duration*	. . . . .
4. *Outcome*	. . . . .
5. *Treatments to be compared*	. . . .
6. *Estimands*	. . . . .
7. *Analysis plan*	. . . . .

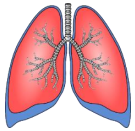
Step 3: The Emulated Trial

The protocol is then adapted to reflect the data:

1. *Eligibility criteria*	. . . . .
2. *Recruitment period*	. . . . .
3. *Follow-up duration*	. . . . .
4. *Outcome*	. . . . .
5. *Treatments to be compared*	. . . .
6. *Estimands*	. . . . .
7. *Analysis plan*	. . . . .

An Application

- **The setting**
  - There are 30,000 hospital admissions for respiratory infections in infants in England every year

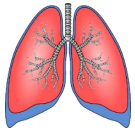


  - Leading cause: seasonal respiratory syncytial virus (RSV) infections, but there is no vaccine against RSV
  - In the UK, immunization with Palivizumab (during RSV season) of *high risk infants*† is recommended, although there is no experimental evidence of efficacy in this population

†premature infants and infants with some chronic heart or lung conditions

An Application

- **The setting**
  - There are 30,000 hospital admissions for respiratory infections in infants in England every year



  - Leading cause: seasonal respiratory syncytial virus (RSV) infections, but there is no vaccine against RSV
  - In the UK, immunization with Palivizumab (during RSV season) of *high risk infants*† is recommended, although there is no experimental evidence of efficacy in this population

The question

Is Palivizumab effective in preventing hospitalization in this high risk population?

† premature infants and infants with some chronic heart or lung conditions

Example

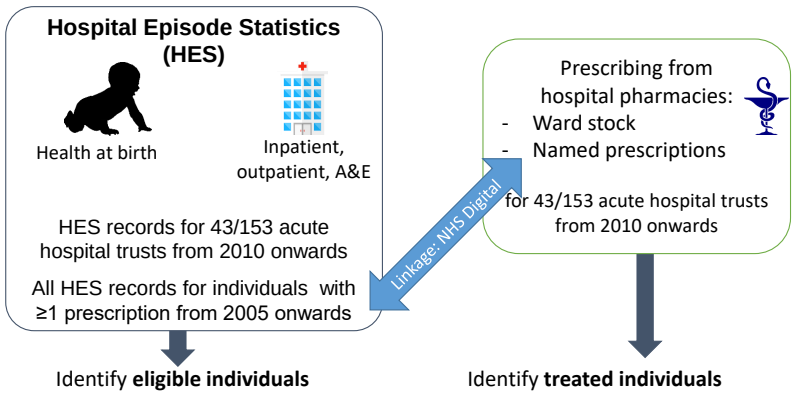
- **The data**:

  Linked data on infants born 2010-2016 [Zylbersztejn *et al.* (2020)]:

  - hospital episode statistics: to identify 'at risk' population and recover hospitalization history
  - hospital prescription records: to recover immunization history

Data: Hospital Treatment Insights (HTI)

- Electronic health records maintained by IQVIA (https://www.iqvia.com/)



More details on HTI data: Rockenschaub, P., Ansell, D., & Shallcross, L. (2017). Linking individual-level data on diagnoses and dispensing for research on antibiotic use: Evaluation of a novel data source from English secondary care. *Pharmacoepidemiology and Drug Safety*, (September), 1–7. https://doi.org/10.1002/pds.4367

Protocol for the Ideal Target Trial

1.	*Eligibility criteria*	"High-risk" infants born in England
2.	*Recruitment period*	2010-2016
3.	*Follow-up duration*	- Start: 1st Oct (of YOB), or DOB if born later - End: the first of hospitalization, end of RSV season, 1st birthday, death
4.	*Outcome*	Hospitalization during RSV season and before 1st birthday
5.	*Treatments to be compared*	1. Start of Pali during RSV season 2. No Pali during the RSV season
6.	*Causal contrasts*	Risk difference
7.	*Analysis plan*	Generalised linear regression model with robust SEs

Protocol for the Ideal Target Trial

1. *Eligibility criteria*	"High-risk" infants born in England
2. *Recruitment period*	2010-2016
3. *Follow-up duration*	- Start: 1st Oct (of YOB), or DOB if born later - End: the first of hospitalization, end of RSV season, 1st birthday, death
4. *Outcome*	Hospitalization during RSV season and before 1st birthday
5. *Treatments to be compared*	1. Start of Pali during RSV season 2. No Pali during the RSV season
6. *Causal contrasts*	Risk difference
7. *Analysis plan*	Generalised linear regression model with robust SEs

Protocol for the Ideal Target Trial

1. *Eligibility criteria*	"High-risk" infants born in England
2. *Recruitment period*	2010-2016
3. *Follow-up duration*	- Start: 1st Oct (of YOB), or DOB if born later - End: the first of hospitalization, end of RSV season, 1st birthday, death
4. *Outcome*	Hospitalization during RSV season and before 1st birthday
5. *Treatments to be compared*	1. Start of Pali during RSV season 2. No Pali during the RSV season
6. *Causal contrasts*	Risk difference
7. *Analysis plan*	Generalised linear regression model with robust SEs

Protocol for the Ideal Target Trial

1. *Eligibility criteria*	"High-risk" infants born in England
2. *Recruitment period*	2010-2016
3. *Follow-up duration*	- Start: 1st Oct (of YOB), or DOB if born later - End: the first of hospitalization, end of RSV season, 1st birthday, death
4. *Outcome*	Hospitalization during RSV season and before 1st birthday
5. *Treatments to be compared*	1. Start of Pali during RSV season 2. No Pali during the RSV season
6. *Causal contrasts*	Risk difference
7. *Analysis plan*	Generalised linear regression model with robust SEs

Protocol for the Ideal Target Trial

1. *Eligibility criteria*	"High-risk" infants born in England
2. *Recruitment period*	2010-2016
3. *Follow-up duration*	- Start: 1st Oct (of YOB), or DOB if born later - End: the first of hospitalization, end of RSV season, 1st birthday, death
4. *Outcome*	Hospitalization during RSV season and before 1st birthday
5. *Treatments to be compared*	1. Start of Pali during RSV season 2. No Pali during the RSV season
6. *Causal contrasts*	Risk difference
7. *Analysis plan*	Generalised linear regression model with robust SEs

Protocol for the Ideal Target Trial

1. *Eligibility criteria*	"High-risk" infants born in England
2. *Recruitment period*	2010-2016
3. *Follow-up duration*	- Start: 1st Oct (of YOB), or DOB if born later - End: the first of hospitalization, end of RSV season, 1st birthday, death
4. *Outcome*	Hospitalization during RSV season and before 1st birthday
5. *Treatments to be compared*	1. Start of Pali during RSV season 2. No Pali during the RSV season
6. *Causal contrasts*	Risk difference
7. *Analysis plan*	Generalised linear regression model with robust SEs

Protocol for the Emulated Trial

1. *Eligibility criteria*	High risk infants born in England for whom sufficient data available‡
2. *Recruitment period*	
3. *Follow-up duration*	
4. *Outcome*	
5. *Treatments to be compared*	
6. *Causal contrasts*	
7. *Analysis plan*	

‡To establish eligibility, treatment, outcome

Protocol for the Emulated Trial

1. *Eligibility criteria*	High risk infants born in England for whom sufficient data available‡
2. *Recruitment period*	
3. *Follow-up duration*	
4. *Outcome*	
5. *Treatments to be compared*	
6. *Causal contrasts*	
7. *Analysis plan*	

‡To establish eligibility, treatment, outcome BUT missing data may lead to selection bias

Protocol for the Emulated Trial

1. *Eligibility criteria*	High risk infants born in England for whom sufficient data available[§]
2. *Recruitment period*	Same
3. *Follow-up duration*	Same
4. *Outcome*	
5. *Treatments to be compared*	
6. *Causal contrasts*	
7. *Analysis plan*	

[§] To establish eligibility, treatment, outcome BUT missing data may lead to selection bias

Protocol for the Emulated Trial

1. *Eligibility criteria*	High risk infants born in England for whom sufficient data available[§]
2. *Recruitment period*	Same
3. *Follow-up duration*	Same
4. *Outcome*	Same but may be affected by misclassification
5. *Treatments to be compared*	
6. *Causal contrasts*	
7. *Analysis plan*	

[§] To establish eligibility, treatment, outcome BUT missing data may lead to selection bias

Protocol for the Emulated Trial

1. *Eligibility criteria*	High risk infants born in England for whom sufficient data available¶
2. *Recruitment period*	Same
3. *Follow-up duration*	Same
4. *Outcome*	Same BUT may be affected by misclassification
5. *Treatments to be compared*	"First recorded use of Pali"
6. *Causal contrasts*	
7. *Analysis plan*	

¶To establish eligibility, treatment, outcome BUT missing data may lead to selection bias

Protocol for the Emulated Trial

1. *Eligibility criteria*	High risk infants born in England for whom sufficient data available‖
2. *Recruitment period*	Same
3. *Follow-up duration*	Same
4. *Outcome*	Same
5. *Treatments to be compared*	"First recorded use of Pali"
6. *Causal contrasts*	Same
7. *Analysis plan*	

‖To establish eligibility, treatment, outcome BUT missing data may lead to selection bias

Protocol for the Emulated Trial

1. *Eligibility criteria*	High risk infants born in England for whom sufficient data available‖
2. *Recruitment period*	Same
3. *Follow-up duration*	Same
4. *Outcome*	Same
5. *Treatments to be compared*	"First recorded use of Pali"
6. *Causal contrasts*	Same
7. *Analysis plan*	Same BUT must control for confounding

‖To establish eligibility, treatment, outcome BUT missing data may lead to selection bias

Controlling for Confounding

Two approaches:

1. Methods that require adjustment for confounders:
   - Stratification/regression,
   - matching, propensity scores
   - G-methods: standardization/g-formula, g-estimation, IP weighting
2. Methods that exploit sources of randomness in the data:
   - instrumental variables (*e.g.* genetic variants)
   - regression discontinuity (*e.g.* randomly varying policy implementation)

Controlling for Confounding

Two approaches:

1. Methods that require adjustment for confounders:
   - Stratification/regression,
   - matching, propensity scores
   - G-methods: standardization/g-formula, g-estimation, IP weighting
2. Methods that exploit sources of randomness in the data:
   - instrumental variables (*e.g.* genetic variants)
   - regression discontinuity (*e.g.* randomly varying policy implementation)

Controlling for Confounding

Two approaches:

1. Methods that require adjustment for confounders:
   - Stratification/regression,
   - matching, propensity scores
   - G-methods: standardization/g-formula, g-estimation, IP weighting
2. Methods that exploit sources of randomness in the data:
   - instrumental variables (*e.g.* genetic variants)
   - regression discontinuity (*e.g.* randomly varying policy implementation)

Both approaches rely on unverifiable assumptions

Controlling for Confounding

Two approaches:

1. Methods that require adjustment for confounders:
   - Stratification/regression,
   - matching, propensity scores
   - G-methods: standardization/g-formula, g-estimation, IP weighting
2. Methods that exploit sources of randomness in the data:
   - instrumental variables (*e.g.* genetic variants)
   - regression discontinuity (*e.g.* randomly varying policy implementation)

To be discussed in the next lecture

Summary

- Addressing causal questions using observational data requires dealing with multiple possible sources of bias.
- Some biases can be avoided by adopting the TTE framework, as this gives formality to the design and analysis steps.
- There are some remaining challenges, however.
- Overall requirements for causal inference:
  (a) a well-defined causal question
  (b) high quality data
  (c) clear study design
  (d) valid data analysis

Summary

- Addressing causal questions using observational data requires dealing with multiple possible sources of bias.
- Some biases can be avoided by adopting the TTE framework, as this gives formality to the design and analysis steps.
- There are some remaining challenges, however.
- Overall requirements for causal inference:

(a) a well-defined causal question
(b) high quality data
(c) clear study design
(d) valid data analysis

Summary

- Addressing causal questions using observational data requires dealing with multiple possible sources of bias.
- Some biases can be avoided by adopting the TTE framework, as this gives formality to the design and analysis steps.
- There are some remaining challenges, however.
- Overall requirements for causal inference:
  (a) a well-defined causal question
  (b) high quality data
  (c) clear study design
  (d) valid data analysis

Summary

- Addressing causal questions using observational data requires dealing with multiple possible sources of bias.
- Some biases can be avoided by adopting the TTE framework, as this gives formality to the design and analysis steps.
- There are some remaining challenges, however.
- Overall requirements for causal inference:

(a) a well-defined causal question
(b) high quality data
(c) clear study design
(d) valid data analysis

Further Reading

- Grodstein *et al.* Hormone therapy and coronary heart disease: the role of time since menopause and age at hormone initiation. *J Women's Health*. 2006;15(1):35–44
- Hernàn *et al.* Specifying a target trial prevents immortal time bias and other self-inflicted injuries in observational analyses *Journal of Clinical Epidemiology* 2016: 79 (2016) 70e75
- Hernàn *et al.* Observational Studies Analyzed Like Randomized Experiments *Epidemiology* 2008;19: 766–779
- Hernàn and Hernandez-Diaz. Beyond the intention to treat in comparative effectiveness research. *Clin Trials* 2012; 9, 48–55
- Hernàn and Robins. Using Big Data to Emulate a Target Trial When a Randomized Trial Is Not Available. *American Journal of Epidemiology*, 2016, 183, 758–764
- Suissa. Immortal time bias in observational studies of drug effects. *Pharmacoepidemiol Drug Saf* 2007; 241–9
- Zylbersztejn *et al.* Access to palivizumab among children at high risk of respiratory syncytial virus complication in English hospitals. *Br J Clin Pharmacol* 2021: 1-12.