Target trial emulation: opportunities and challenges

Barbra Dickerman, PhD

Assistant Professor, Zhu Family Center for Global Cancer Prevention

Disclosures

➤ This work was supported by
  - National Institutes of Health National Cancer Institute (R00 CA248335)
  - U.S. Department of Veterans Affairs Office of Research and Development Cooperative Studies Program Epidemiology Center at the VA Boston Healthcare System through CSP #2032

➤ The contents of this presentation reflect the views of our study team and not necessarily the position of the VA or U.S. Government
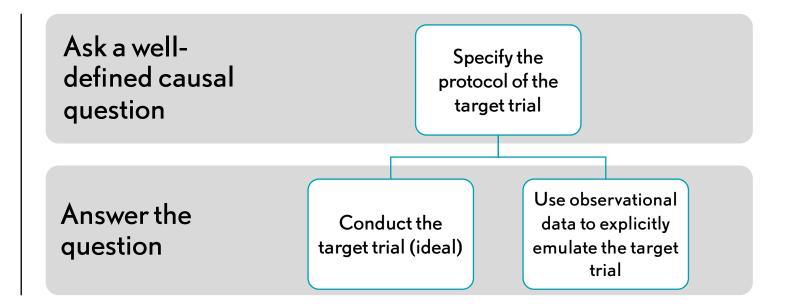
The target trial

↘ The (hypothetical) randomized trial that would answer the causal question of interest

Comparative effectiveness or safety research

Ask a well-defined causal question	Specify the protocol of the target trial
Answer the question	Conduct the target trial (ideal) / Use observational data to explicitly emulate the target trial

Target trial protocol components

- Eligibility criteria
- Treatment strategies
- Treatment assignment
- Follow-up
- Outcome(s)
- Causal contrast(s)
- Analysis plan

An observational study needs to explicitly emulate all of these

Why does this matter?

✎ Why do we want to explicitly emulate a target trial when using observational data for causal inference?

1. Reduces bias in observational analyses
2. Generates actionable evidence to inform decision-making

Let's look at some examples...

Reducing bias in observational analyses
Case #1 : Statins and cancer

- Randomized-observational discrepancies
- Demonstrated that these (1) appear to be due to analytic flaws in the observational studies and not any inherent problems with the observational data, and (2) disappear when observational data are analyzed using methods consistent with the target trial framework
  - Dickerman et al, *Nature Medicine* 2019

nature medicine

ANALYSIS

https://doi.org/10.1038/s41591-019-0597-x

Avoidable flaws in observational analyses: an application to statins and cancer

Barbra A. Dickerman [ID][1*], Xabier García-Albéniz[1,2], Roger W. Logan[1], Spiros Denaxas[3,4,5] and Miguel A. Hernán[1,6,7]

Generating timely, actionable evidence
Case #2: Covid-19 vaccines

- Led the first study to emulate a target trial of the **comparative effectiveness** of mRNA-based Covid-19 vaccines

The NEW ENGLAND JOURNAL of MEDICINE

ESTABLISHED IN 1812   JANUARY 13, 2022   VOL. 386   NO. 2

Comparative Effectiveness of BNT162b2 and mRNA-1273 Vaccines in U.S. Veterans

Barbra A. Dickerman, Ph.D., Hanna Gerlovin, Ph.D., Arin L. Madenci, M.D., Ph.D., Katherine E. Kurgansky, M.P.H., Brian R. Ferolito, M.Sc., Michael J. Figueroa Muñiz, B.Sc., David R. Gagnon, M.D., Ph.D., M.P.H., J. Michael Gaziano, M.D., M.P.H., Kelly Cho, Ph.D., Juan P. Casas, M.D., Ph.D., and Miguel A. Hernán, M.D., Dr.P.H.

Generating timely, actionable evidence
Case #2: Covid-19 vaccines

✎ Led additional studies that provided evidence for the **comparative safety** of these vaccines

Research

JAMA Internal Medicine | Original Investigation

Comparative Safety of BNT162b2 and mRNA-1273 Vaccines in a Nationwide Cohort of US Veterans

Barbra A. Dickerman, PhD; Arin L. Madenci, MD, PhD; Hanna Gerlovin, PhD; Katherine E. Kurgansky, MPH; Jessica K. Wise, MPH; Michael J. Figueroa Muñiz, BSc; Brian R. Ferolito, MSc; David R. Gagnon, MD, PhD, MPH; J. Michael Gaziano, MD, MPH; Kelly Cho, PhD, MPH; Juan P. Casas, MD, PhD; Miguel A. Hernán, MD, DrPH

Generating timely, actionable evidence
Case #2: Covid-19 vaccines

✎ As well as the **comparative effectiveness of booster doses**

nature microbiology

Comparative effectiveness of third doses of mRNA-based COVID-19 vaccines in US veterans

Barbra A. Dickerman [1,2,9], Hanna Gerlovin [3,9] ✉, Arin L. Madenci [1,2,4], Michael J. Figueroa Muñiz [3,5], Jessica K. Wise[3], Nimish Adhikari[3,5], Brian R. Ferolito[3], Katherine E. Kurgansky[3,6], David R. Gagnon [3,5], Kelly Cho[3,7], Juan P. Casas[3,7] & Miguel A. Hernán [1,2,8]

Challenges in target trial emulation

1) Identifying an appropriate data source
2) Feature extraction / phenotyping
3) Dataset assembly
4) Analytic pipelines

Challenges in target trial emulation

1) Identifying an appropriate data source

- Key requirement for successful emulation: well-defined causal question is paired with a **sufficiently rich data source** to answer it

- Previous "failures" of target trial emulation can be explained by inadequate pairings and not any inherent problems with the methodological approach itself
  - E.g., estimating the effect of preventive services on mortality using administrative claims data
  - García-Albéniz et al., AJE 2019

Challenges in target trial emulation

1) Identifying an appropriate data source

2) Feature extraction / phenotyping

- Even if information on key confounders is available, extracting it can be challenging

- In our work:
  - Leverage existing phenotype libraries
  - Integrate structured & unstructured data
    - E.g., using ML/AI to extract information from clinical notes
  - Incorporate expertise in clinical domain & observational database
    - To identify the key variables, including confounders
    - To understand where relevant information is recorded – and how, when, for whom

Challenges in target trial emulation

1) Identifying an appropriate data source
2) Feature extraction / phenotyping

3) Dataset assembly

- Requires joining many features into a longitudinal dataset
- In our work:
  - Design data pipelines to systematically construct analytic datasets
  - Incorporate expertise in the clinical domain, observational database, & computationally efficient programming
    - To ensure the time resolution of the data reflects the decision-making process
    - To understand where relevant features are recorded in the database, how often they are refreshed, and how to efficiently join them

Challenges in target trial emulation

1) Identifying an appropriate data source
2) Feature extraction / phenotyping
3) Dataset assembly

4) Analytic pipelines

- Application of state-of-the-art causal methods to large health databases can be complex, computationally intensive
- In our work:
  - Develop flexible toolkits with detailed documentation
  - Incorporate expertise in the clinical domain, causal inference methodology, and computationally efficient programming
    - To design analyses to investigate potential biases of concern
    - To efficiently work within the constraints of available computational resources

Our best chance at unlocking the full potential of real-world data to support health decisions

- **Is in the combination of**
  - High-quality data
    - Integrated and maintained large health care databases
  - Experts
    - In that health data, and in using it for causal inference
- **The target trial framework can be a useful tool to**
  - Avoid the avoidable biases in observational analyses
  - Generate timely, actionable evidence to inform decision-making

Acknowledgements

Members of the CAUSALab and VA-CAUSAL Methods Core

barbra_dickerman@g.harvard.edu

CAUSALab
A Center to Learn What Works

HARVARD
T.H. CHAN
SCHOOL OF PUBLIC HEALTH