From healthcare data to decisions: the target trial framework

Barbra Dickerman, PhD

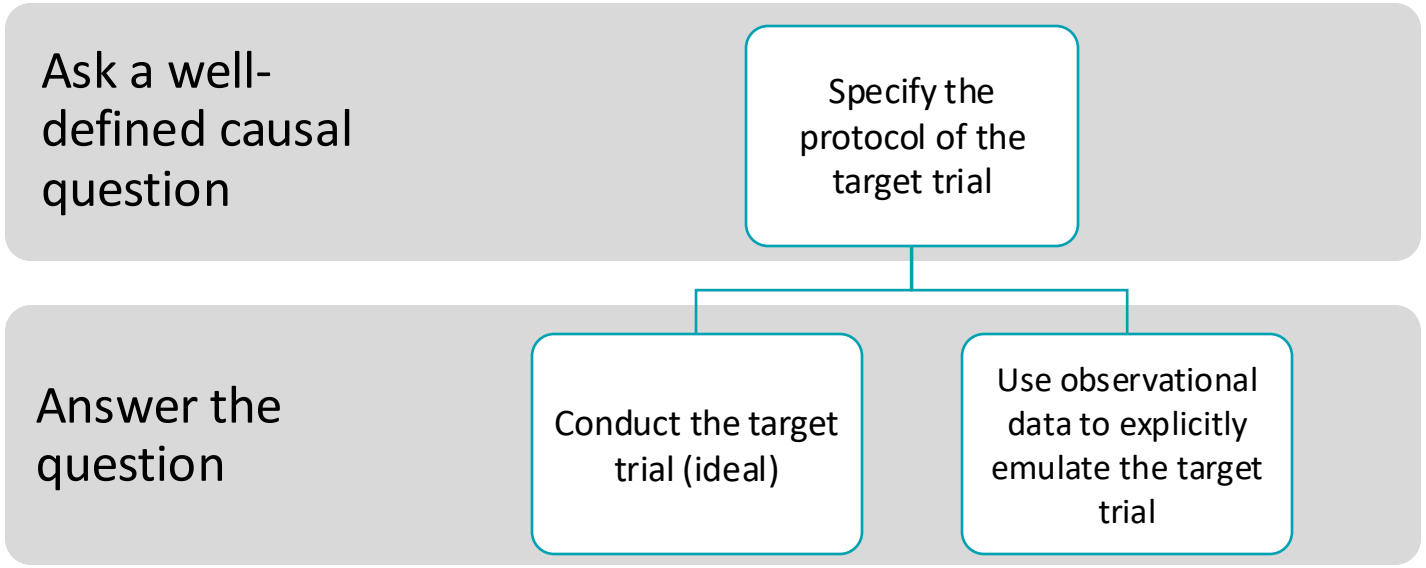
Assistant Professor, Zhu Family Center for Global Cancer Prevention

HARVARD
T.H. CHAN
SCHOOL OF PUBLIC HEALTH

The target trial

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

Comparative effectiveness or safety research



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?

- **Because it reduces bias**
  - And can help to resolve discrepancies between effect estimates from observational studies and randomized trials
  - Hernán & Robins, *Am J Epidemiol* 2016

Let's look at an example...

Do statins prevent cancer?

- Observational studies: association between statins and lower cancer risk
  - Some reported an implausible 50-65% lower risk
- Meta-analyses of randomized trials: no effect
- Why the discrepancy?
  - Confounding due to lack of randomization in the observational studies?
    - Unlikely because cancer is not an intended effect of treatment
  - Analytic flaws in the observational studies?

Do statins prevent cancer?

- Demonstrated that these randomized-observational discrepancies (1) appear to be due to the analytic approach ("self-inflicted biases") 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

Protocol of a target trial of statin therapy and cancer incidence

- Eligibility criteria
- Treatment strategies
- Assignment procedures
- Outcomes
- Follow-up
- Causal contrasts
- Statistical analysis

Observational data available for emulation:

Nationwide database of EHRs in the UK

- Clinical Practice Research Database (CPRD) primary care data
- Linked to hospital, death registries
- Accessed through the CALIBER resource

Protocol of a target trial of statin therapy and cancer incidence

Eligibility criteria

Treatment strategies

Assignment procedures

Outcomes

Follow-up

Causal contrasts

Statistical analysis

SPECIFICATION

- Individuals aged ≥30 years between January 1998-February 2016
- No history of cancer
- No statin prescription in the past year
- No statin contraindication (hepatic impairment, myopathy)
- LDL cholesterol <5 mmol/L
- ≥1 year of up-to-standard data in a CPRD practice

EMULATION

Same as for the target trial

Protocol of a target trial of statin therapy and cancer incidence

Eligibility criteria

Treatment strategies

Assignment procedures

Outcomes

Follow-up

Causal contrasts

Statistical analysis

SPECIFICATION

1. Initiation of any statin therapy at baseline and continuation over follow-up until the development of a contraindication, or
2. No initiation of statin therapy over follow-up until the development of an indication

EMULATION

Same as for the target trial.

Defined the date of medication initiation to be the first date of a prescription.

Calculated discontinuation dates using the daily dose and quantity of pills in the prescription.

Protocol of a target trial of statin therapy and cancer incidence

Eligibility criteria

Treatment strategies

Assignment procedures

Outcomes

Follow-up

Causal contrasts

Statistical analysis

SPECIFICATION

Individuals are randomly assigned to a strategy at baseline and are aware of the strategy to which they have been assigned

EMULATION

Classified individuals into 1 of 2 treatment groups (statin initiators or non-initiators) based on the strategy that their data were compatible with at baseline.

Assumed randomization conditional on baseline variables.

Protocol of a target trial of statin therapy and cancer incidence

Eligibility criteria

Treatment strategies

Assignment procedures

Outcomes

Follow-up

Causal contrasts

Statistical analysis

SPECIFICATION

Total cancer and 7 site-specific invasive cancers: female breast, colorectal, hematological, melanoma, lung, prostate, urothelial

EMULATION

Same as for the target trial

Protocol of a target trial of statin therapy and cancer incidence

Eligibility criteria

Treatment strategies

Assignment procedures

Outcomes

Follow-up

Causal contrasts

Statistical analysis

SPECIFICATION

From the day of treatment assignment (baseline) until earliest of:

- First cancer diagnosis
- Death
- Loss to follow-up
- 10 years after baseline
- End of study period (February 2016)

EMULATION

Same as for the target trial

Protocol of a target trial of statin therapy and cancer incidence

Eligibility criteria

Treatment strategies

Assignment procedures

Outcomes

Follow-up

Causal contrasts

Statistical analysis

SPECIFICATION

Intention-to-treat effect

Per-protocol effect

EMULATION

Observational analogues of the intention-to-treat and per-protocol effects

Protocol of a target trial of statin therapy and cancer incidence

Eligibility criteria

Treatment strategies

Assignment procedures

Outcomes

Follow-up

Causal contrasts

Statistical analysis

SPECIFICATION

Pooled logistic regression to estimate standardized risk curves and hazard ratios.

Intention-to-treat analysis: apply inverse-probability (IP) weights to adjust for pre- and post-baseline prognostic factors associated with loss to follow-up.

Per-protocol analysis: censor individuals if/when they deviate from their assigned treatment strategy and apply IP weights to adjust for pre- and post-baseline prognostic factors associated with adherence and loss to follow-up.

G-methods

EMULATION

Same as for the target trial with sequential emulation and adjustment for baseline covariates

Intention-to-treat analysis

- Hazard ratio

$$\text{logit}\left(\Pr[Y_{t+1} = 1 | A_0, L_0, \bar{Y}_t = 0]\right) = \beta_{0,t} + \beta_1 A_0 + \beta_2{}^T L_0$$

- Absolute risks

$$\text{logit}\left(\Pr[Y_{t+1} = 1 | A_0, L_0, \bar{Y}_t = 0]\right) = \beta_{0,t} + \beta_1 A_0 + \beta_2{}^T L_0 + \beta_3 A_0 t$$

Y_{t+1}	Indicator for outcome of interest at month t
$\beta_{0,t}$	Time-varying intercept, estimated as a constant plus linear and quadratic terms for follow-up month t
A_0	Indicator for assigned treatment strategy
L_0	Vector of potential baseline confounders for each individual

Per-protocol analysis

$$W_t^A = \prod_{k=0}^{t} \frac{1}{f(A_k|\bar{A}_{k-1}, \bar{L}_k, \bar{Y}_{k-1} = 0)}$$

- Hazard ratio

$$\text{logit}\left(\Pr[Y_{t+1} = 1|A_0, L_0, \bar{Y}_t = 0, \bar{C}_{t+1} = 0]\right) = \beta_{0,t} + \beta_1 A_0 + \beta_2{}^T L_0$$

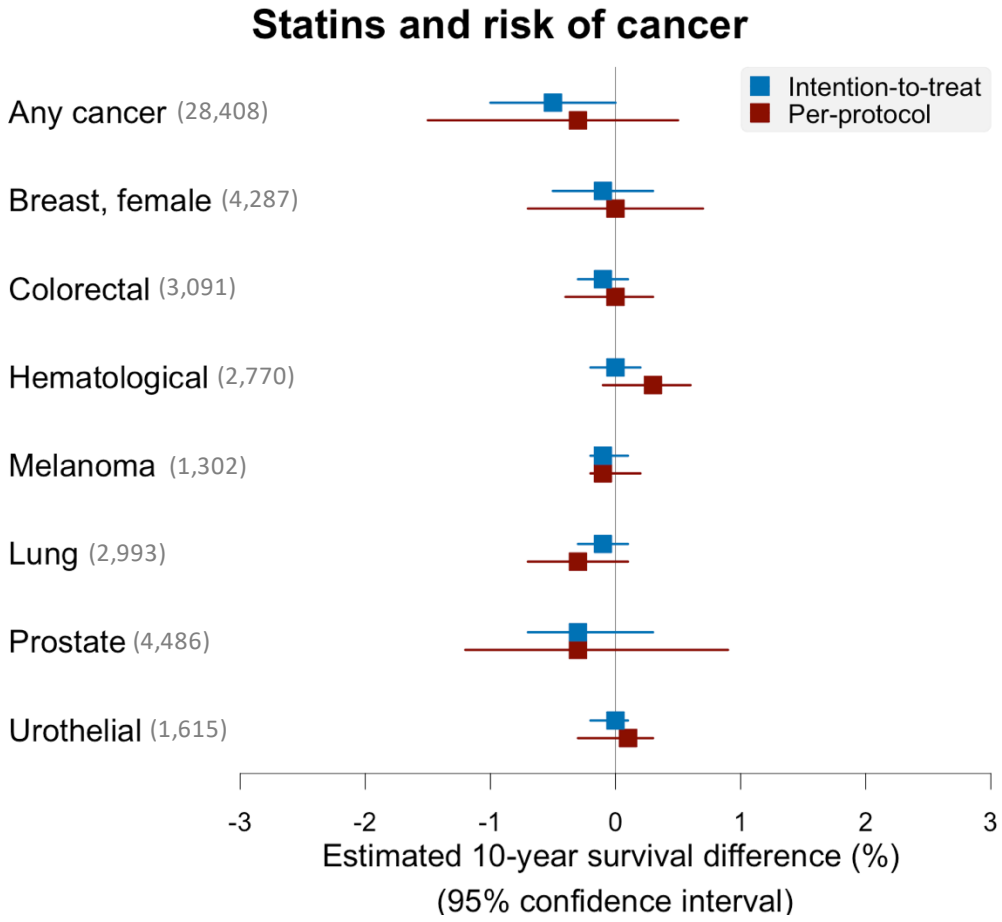
- Absolute risks

$$\text{logit}\left(\Pr[Y_{t+1} = 1|A_0, L_0, \bar{Y}_t = 0, \bar{C}_{t+1} = 0]\right) = \beta_{0,t} + \beta_1 A_0 + \beta_2{}^T L_0 + \beta_3 A_0 t$$

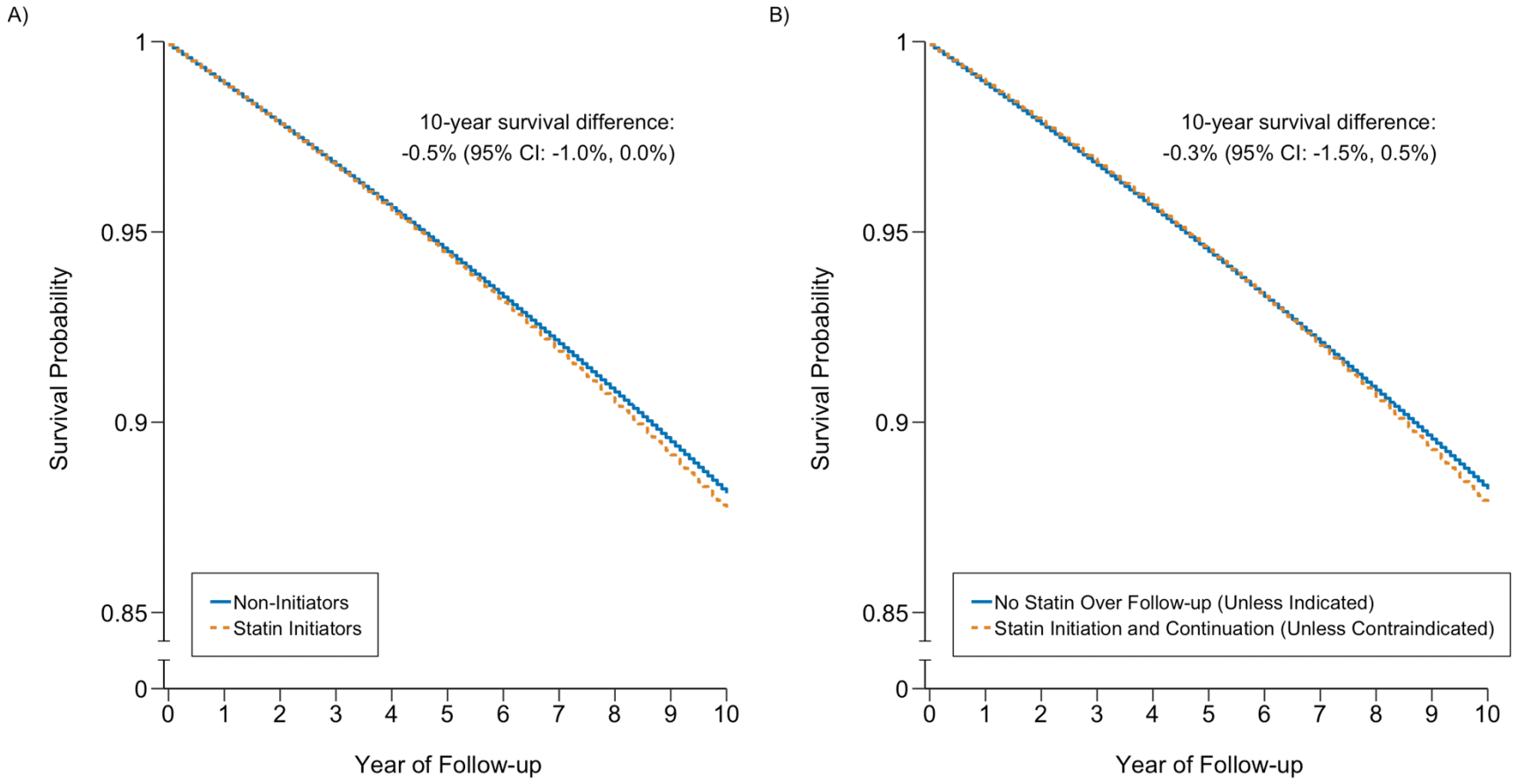
Y_{t+1}	Indicator for outcome of interest at month t
$\beta_{0,t}$	Time-varying intercept, estimated as a constant plus linear and quadratic terms for follow-up month t
A_0	Indicator for assigned treatment strategy
L_0	Vector of potential baseline confounders for each individual

Statins do not appear to influence cancer incidence over up to 10 years of follow-up

Effect estimates from the target trial emulation were comparable to those from meta-analyses of randomized trials



Standardized cancer-free survival curves under each strategy were overlapping



Previous observational study: OR for lung cancer 0.23 (95% CI: 0.20, 0.26)

- Among long-term statin users (>4 years) vs. nonusers
- 2 key deviations from the target trial:
  - Including prevalent statin users at baseline
    → Selection bias
  - Classifying individuals based on achieved treatment duration
    → Immortal time bias
- When we applied these decisions to our data:
  - HR for lung cancer 0.27 (0.25, 0.29)
  - HR for total cancer 0.23 (0.22, 0.24)

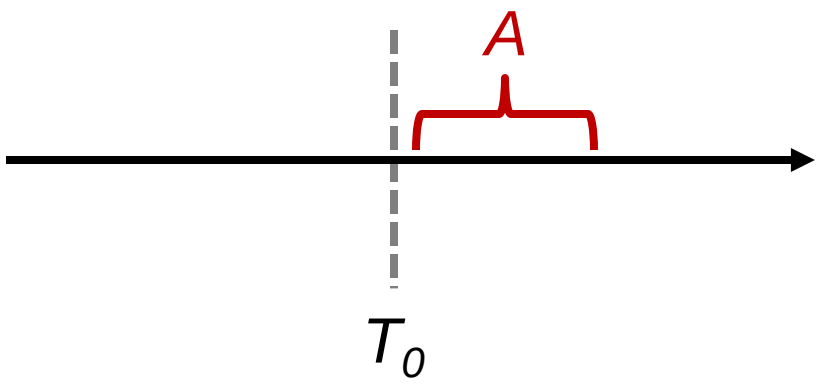
Here, the discrepancy between observational-randomized evidence was not due to confounding

- Critics of observational studies often focus on confounding from lack of randomization
  - Can be minimized, but not eliminated
- Less attention: self-inflicted biases due to **study design flaws**
  - Easily prevented

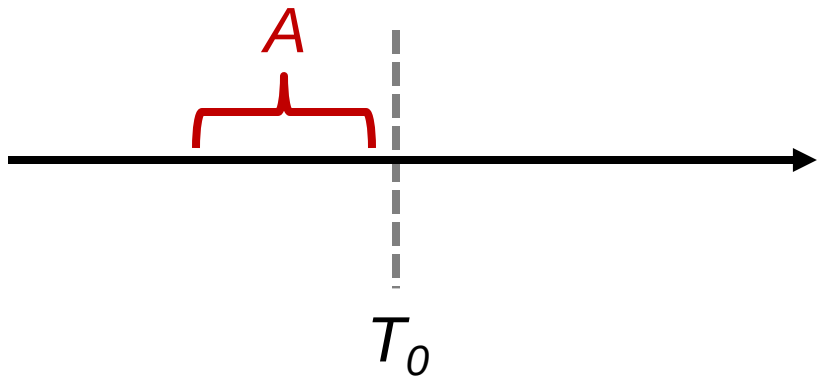
A basic tenet of any trial is the alignment of eligibility and treatment assignment with time zero

Some observational analyses **unhitch treatment assignment from time zero** by

Classifying individuals based on achieved treatment duration



Including prevalent users at baseline



Explicitly emulating a target trial

- Ensures synchronization of time zero, eligibility ascertainment, and treatment assignment
- Avoids the avoidable biases
  - Immortal time bias and other design-related biases
- Reduces discrepancies between observational and randomized evidence

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

Acknowledgements

Project Collaborators

- Miguel Hernán
- Xabier García-Albéniz
- Roger Logan
- Spiros Denaxas

Support

- NIH/NCI R00 CA248335
- CANCER-CAUSAL Program
- Zhu Family Center for Global Cancer Prevention

barbra_dickerman@g.harvard.edu

CAUSALab
A Center to Learn What Works

HARVARD
T.H. CHAN
SCHOOL OF PUBLIC HEALTH