StatsVia

INSIGHTS · REAL-WORLD EVIDENCE

Target trial emulation: designing observational studies regulators accept

Specify the randomised trial you would run, then emulate it with real-world data to avoid self-inflicted biases.

StatsVia Insights · statsvia.com · connect@statsvia.com

WHY IT MATTERS

Many observational studies fail because of design, not data

COMMON SELF-INFLICTED BIASES

- **Immortal time bias:** follow-up starts before treatment is assigned, so treated patients get guaranteed event-free time
- **Prevalent-user bias:** including people already on treatment selects those who tolerated it and survived
- **Misaligned time zero:** eligibility, treatment assignment and start of follow-up happen at different times

The idea (Hernán & Robins, 2016): write the protocol of the pragmatic randomised trial that would answer the question, then emulate each element using observational data.

The key rule: eligibility is met, treatment strategy is assigned and follow-up starts at the same moment, the way randomisation aligns them in a trial.

THE PROTOCOL

Seven components: target trial versus emulation

Component	Target trial (what we would do)	Emulation with RWD (how we mimic it)
Eligibility	Adults with type 2 diabetes, no prior use of either drug	Same criteria applied using claims or EHR data before time zero
Treatment strategies	Initiate drug A vs initiate drug B	New users of A or B, identified by first prescription
Assignment	Randomised, open-label	Adjust for baseline confounders (PS weighting or matching)
Follow-up	From randomisation until outcome, death, loss or 5 years	From first prescription (time zero); same end rules
Outcomes	Hospitalisation for heart failure	Validated algorithm using diagnosis codes
Causal contrast	Intention-to-treat and per-protocol effects	Observational analogue of ITT; per-protocol via censoring and IPW
Analysis plan	Cox model, cumulative incidence curves	Weighted Cox model or pooled logistic; adjusted risk curves

DESIGN AND ANALYSIS TOOLKIT

Methods that make the emulation credible

Design

- New-user, active-comparator designs
- Washout and look-back windows for covariates
- Sequential trials when eligibility recurs
- Grace periods for treatment start

Confounding and censoring

- Propensity scores: matching, IPTW, overlap weights
- Clone–censor–weight for sustained strategies
- g-methods: IPW of censoring, parametric g-formula
- TMLE and doubly robust estimators

Robustness

- Negative control outcomes and exposures
- E-values and quantitative bias analysis
- Benchmarking against a known RCT result
- Covariate balance diagnostics (SMD < 0.1)

Feasibility first: check sample size, data completeness and outcome validity before running outcome analyses, and freeze the protocol and SAP before looking at results.

APPLICATIONS

Evidence that well-designed emulations reproduce trial results

POSTMENOPAUSAL HORMONE THERAPY

Reconciling observational data with the WHI trial

Re-analysing the Nurses' Health Study as an emulated trial of initiators, with aligned time zero, produced coronary heart disease estimates consistent with the randomised WHI results (Hernán et al., 2008).

RCT-DUPLICATE

32 trials emulated with insurance claims

When trial design could be closely emulated, database studies largely agreed with the randomised results; differences were often explained by design features that could not be emulated (Wang et al., 2023).

COVID-19 VACCINES

Comparative effectiveness in US veterans

An emulated trial compared BNT162b2 with mRNA-1273 using matched recipients, giving head-to-head evidence no RCT had provided (Dickerman et al., 2022).

REGULATORY USE

External control arms and safety studies

Target trial principles underpin external controls for single-arm trials in rare disease and oncology, and post-authorisation safety studies.

WHAT REGULATORS LOOK FOR

Regulatory expectations and references

EXPECTATIONS

- Protocol and SAP finalised and ideally registered before outcome analysis
- Data shown to be fit for purpose: relevant and reliable
- Transparent feasibility assessment and covariate definitions
- Pre-specified sensitivity and bias analyses
- Reporting per the TARGET guideline, plus STROBE and RECORD
- Early dialogue with FDA, EMA or other agencies on the design

REFERENCES

1. Hernán MA, Robins JM. Using big data to emulate a target trial when a randomized trial is not available. Am J Epidemiol 2016;183:758–764.
2. Hernán MA, Alonso A, Logan R, et al. Observational studies analyzed like randomized experiments: postmenopausal hormone therapy and coronary heart disease. Epidemiology 2008;19:766–779.
3. Wang SV, Schneeweiss S, et al. Emulation of randomized clinical trials with nonrandomized database analyses: results of 32 clinical trials. JAMA 2023;329:1376–1385.
4. Dickerman BA, Gerlovin H, Madenci AL, et al. Comparative effectiveness of BNT162b2 and mRNA-1273 vaccines in U.S. veterans. N Engl J Med 2022;386:105–115.
5. Cashin AG, et al. Transparent reporting of observational studies emulating a target trial: the TARGET guideline. JAMA 2025.
6. FDA. Considerations for the use of real-world data and real-world evidence to support regulatory decision-making for drug and biological products. Guidance, 2023.
7. ICH M14. General principles on plan, design and analysis of pharmacoepidemiological studies that utilize real-world data for safety assessment of medicines. 2025.