AS15

Targeted Maximum Likelihood Estimation: A Unified Approach for Confounding and Selection Bias

Mahjoob O Majeed
Catalyst Flex
mahjoob.majeed@catalystcr.com

Noel S John
Catalyst Flex
noel.john@catalystcr.com

PHUSE APAC CONNECT

February 2026

CATALYST FLEX

Agenda

- Introduction
- Confounding and Selection Bias
- Conventional Method for Bias
- Addressing Bias with TMLE

CATALYST FLEX

Introduction

Goal of clinical and observational research is to estimate the true effect of a treatment or intervention on an outcome of interest.

Why TMLE is need?

This goal is often challenged by various forms of bias, which can distort the relationship between treatment and outcome.

Bias occurs when the estimated effect systematically differs from the true effect, leading to incorrect or misleading conclusions.

Conventional methods don’t always work well, especially when we have lots of data or variables.

Key Focus Areas

Confounding and Selection Bias

Conventional Method for Bias

Addressing Bias with TMLE

CATALYST FLEX

Confounding and Selection Bias

Confounding bias

- An external variable influences both the treatment and the outcome.

  (Example: Smoking is associated with both coffee drinking and heart disease, which can distort the true relationship between coffee and heart disease.)

Selection bias

- When the study population is not representative of the target population due to the way participants are selected or retained in the study.

  (Example: A diabetes drug trial recruits only young, healthy volunteers, excluding older or sicker patients, so results may not apply to the entire diabetic population.)

Researchers should identify and adjust for confounders and ensure inclusive study design with minimal dropout or selection bias.

CATALYST FLEX

Conventional Methods in Clinical Trials

Randomization: Assigning participants randomly to treatment groups ensures that confounders are balanced across groups, minimizing confounding bias.

Matching: Participants with similar characteristics (e.g., age, gender, disease severity) are paired across treatment groups to control for confounders.

Stratification: Dividing participants into subgroups based on confounders and analyzing them separately helps reduce bias.

Inclusive Study Design: Ensuring diverse and representative participant recruitment minimizes selection bias.

Minimizing Dropout: Strategies to retain participants and avoid differential loss prevent bias in follow-up data.

Propensity Score Methods: Techniques like propensity score matching or inverse probability weighting adjust for differences in baseline characteristics between groups, reducing confounding.

CATALYST FLEX

Limitations

These methods reduce bias but are not foolproof.

Randomization can fail in small samples.

Matching and stratification become difficult when there are many confounders.

Propensity score methods can be unstable if models are misspecified.

High-dimensional confounders increase variance and instability in propensity-based methods.

Inclusive recruitment strategies can be difficult to implement in practice.

Retention of participants is especially challenging in large or multi-center trials.

Targeted Maximum Likelihood Estimation (TMLE)

- TMLE addresses these limitations and is well-suited for complex trial designs and real-world clinical data, where confounding and selection bias are often unavoidable.
- Integrates two models simultaneously:
  - Outcome regression model
  - Treatment assignment (propensity score) model
- Directly targets the causal parameter of interest (e.g., average treatment effect)
- Provides double robustness

Core Components of TMLE

1. Outcome Regression Model (Q-model)

This model estimates the expected outcome given treatment and covariates:

$$Q(A,W)=E(Y|A,W)$$

Where Y: outcome, A: treatment or exposure, W: baseline covariates.

This model captures how outcomes depend on treatment after adjusting for confounders.

2. Treatment Assignment Model (Propensity Score Model, g-model)

This model estimates the probability of receiving treatment given covariates:

$$g(W)=P(A=1|W)$$

Where Y: outcome, A: treatment or exposure, W: baseline covariates.

This model accounts for confounding and corrects the imbalance in treatment assignment.

CATALYST FLEX

Targeting the Causal Parameter

- Unlike conventional methods that estimate nuisance parameters (e.g., regression coefficients), TMLE directly targets the causal estimand (such as ATE):

$$ATE = E[Y(1) - Y(0)]$$

- The targeting step ensures that:
  - The estimator satisfies the **efficient influence function (EIF)**
  - Bias is reduced in the direction that matters for the causal parameter

This makes TMLE fundamentally different from “plug-in” estimators.

TMLE Estimation Steps

1. Start with initial estimates of:
    - Outcome regression $\hat{Q}_0(A, W)$
    - Propensity score $\hat{g}(W)$
2. Update the outcome model using a **clever covariate**, derived from the propensity score:

$$H(A, W) = \frac{A}{\hat{g}(W)} - \frac{1 - A}{1 - \hat{g}(W)}$$

3. Perform a targeted fluctuation (often via logistic or linear regression) that:
    - Adjusts the outcome estimate
    - Leaves the treatment mechanism unchanged

This step “targets” the estimate toward the causal parameter.

CATALYST FLEX

Double Robustness

- Consistent if either outcome regression (Q) OR treatment model (g) is correctly specified
- Does not require both models to be correct simultaneously
- When both Q and g are correctly specified → achieves semiparametric efficiency bound
- Yields optimal efficiency

TMLE in Clinical Research

Unified & Robust: Addresses confounding and selection bias effectively.

Efficient: Achieves optimal performance under correct model specification

Simulation Evidence: Outperforms traditional methods in bias reduction and estimator stability.

Future Ready: Well-suited for complex data structures in modern clinical trials

Powerful Tool: Enables reliable causal inference for advancing clinical research.

CATALYST FLEX

THANK YOU

QUESTIONS?

CATALYST FLEX