Paper AS15

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

Mahjoob O M, Catalyst Flex, Kochi, India
Noel S John, Catalyst Flex, Trivandrum, India

ABSTRACT

This paper focuses on the use of Targeted Maximum Likelihood Estimation (TMLE) as a unified approach to handle confounding and selection bias in clinical trials. Traditional methods such as propensity score matching and inverse probability weighting often fail to address high-dimensional confounders and may lead to biased treatment effect estimates in the presence of selection bias. In contrast, TMLE combines outcome regression and treatment assignment models, targeting the causal parameter directly while offering a doubly-robust approach. This paper highlights the strengths of TMLE in providing efficient and reliable treatment effect estimates, even in complex trial designs where bias is inevitable. Finally, this article shows how TMLE can be applied in clinical research using the tmle R package to model treatment and outcomes. Through simulated data, we highlight how TMLE reduces bias from confounding and selection, resulting in more accurate and reliable results.

INTRODUCTION

In clinical and observational research, the ultimate goal is to estimate the true effect of treatment or intervention on an outcome of interest. However, 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.

Traditional methods such as multivariable regression, propensity score matching (PSM), and inverse probability weighting (IPW) attempt to correct these biases. However, these methods rely heavily on correct model specification and struggle with high-dimensional covariates or nonlinear relationships. In addition, selection bias is often handled separately from confounding, leading to fragmented and potentially inefficient estimation strategies.

Targeted Maximum Likelihood Estimation (TMLE) provides a unified framework that simultaneously addresses confounding and selection bias while directly targeting the causal parameter. By combining outcome modeling and treatment modeling within a principled estimation procedure, TMLE improves robustness and efficiency in complex trial settings.

BIASES ADDRESSED IN THIS PAPER

Confounding bias

Confounding arises when baseline covariates influence both treatment assignment and outcomes. Failure to appropriately adjust for these variables results in biased treatment effect estimates. High-dimensional confounders, interactions, and nonlinear effects further exacerbate this problem.

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

Selection bias

Selection bias occurs when inclusion in the analysis depends on variables related to treatment or outcome, such as informative dropout or missing outcomes. Methods such as IPW can adjust for selection bias but are often unstable when estimated probabilities are extreme.

(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.)

When confounding and selection bias occur simultaneously, causal effect estimation becomes particularly challenging. Confounding distorts the association between treatment and outcome due to systematic differences in baseline characteristics, while selection bias further alters the observed data by restricting analysis to a non-representative subset of the study population. Together, these biases compound one another: treatment groups may differ in unbalanced covariate distributions, and the observed outcomes may no longer reflect the target population due to outcome-dependent missingness or informative censoring.

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

CONVENTIONAL METHODS IN CLINICAL TRIALS

Clinical trials rely on well-established methodological approaches to minimize confounding and selection bias, thereby strengthening the validity of causal conclusions. These methods aim to ensure that observed differences in outcomes between treatment groups are attributable to the intervention itself rather than to systematic differences in patient characteristics.

Some commonly used methods to reduce confounding and selection bias in clinical trials:

Randomization

Randomization assigns patients to treatment arms by chance, ensuring that clinical characteristics such as disease severity, comorbidities, and demographic factors are evenly distributed between groups. This balance allows clinicians to attribute differences in outcomes directly to the intervention rather than to pre-existing patient differences.

Under perfect adherence and complete follow-up, randomization yields unbiased estimators of causal treatment effects.

Matching

Matching pairs patients in different treatment groups who share similar clinical profiles, such as age, sex, baseline disease stage, or risk factors. This approach creates clinically comparable groups and is particularly useful when randomization is not feasible. By restricting comparisons to matched individuals, this approach improves covariate balance but may reduce efficiency and induce selection bias if unmatched units are systematically excluded.

Stratification

Stratification divides patients into clinically meaningful subgroups based on important prognostic factors, such as disease severity or biomarker status. Treatment effects are assessed within each subgroup, ensuring comparisons are made among patients with similar clinical risk. This method is effective for a limited number of strong clinical predictors but becomes less practical as patient complexity increases.

Inclusive Study Design

An inclusive study design seeks to enroll patients who reflect the diversity of real-world clinical populations. Broad eligibility criteria and equitable recruitment reduce selection bias and improve the generalizability of trial findings, ensuring that results are applicable across different patient subgroups encountered in routine clinical practice.

Minimizing Dropout

Patient retention is critical for maintaining unbiased outcome assessment. Differential dropouts such as higher withdrawal among sicker patients can distort treatment comparisons. Minimizing dropout reduces dependence on missing-at-random assumptions and decreases the need for complex modeling of selection mechanisms.

Propensity Score Methods

Propensity score methods adjust for differences in baseline clinical characteristics by estimating each patient's likelihood of receiving a given treatment. Techniques such as matching, stratification, and inverse probability weighting use this score to achieve covariate balance. However, these methods rely on correct specification of the propensity model and the assumption of no unmeasured confounding.

Inverse Probability Weighting (IPW)

Inverse probability weighting is a specific propensity score–based technique that creates a weighted pseudo-population in which treatment assignment is independent of baseline covariates. By weighting individuals inversely to their probability of receiving the treatment they actually received, IPW allows estimation of marginal treatment effects for the target population. While IPW efficiently uses all available data and directly targets population-level estimands, it is sensitive to extreme propensity scores and can yield unstable estimates when treatment groups have limited overlap.

LIMITATIONS

Despite their widespread adoption, conventional methods for controlling confounding and selection bias in clinical trials have notable limitations that can compromise causal inference. Randomization, while theoretically ensuring balance in both measured and unmeasured covariates, may fail to achieve adequate balance in small or moderately sized trials, leading to chance imbalances in important prognostic factors. Matching and stratification improve comparability between treatment groups but become increasingly difficult to implement as the number of confounders grows, often resulting in loss of sample size, reduced statistical efficiency, and residual imbalance when patient characteristics are complex or high-dimensional.

Propensity score–based methods, including inverse probability weighting (IPW), offer flexible tools for adjusting for multiple baseline covariates but are highly sensitive to the correct specification of the propensity score model. Misspecification can lead to inadequate covariate balance and biased treatment effect estimates. In addition, IPW can produce extreme weights when some patients have very high or very low probabilities of receiving a particular treatment, inflating variance and making estimates unstable, especially in settings with limited overlap between

treatment groups. Like all propensity score methods, IPW also relies on the strong assumption that all relevant confounders are measured, which is often difficult to guarantee in clinical research.

Inclusive recruitment and long-term participant retention, while essential for external validity and unbiased outcome assessment, present practical challenges, particularly in large, multi-center trials where logistical, socioeconomic, and clinical factors may differentially affect enrollment and follow-up. Regression-based approaches, commonly used to adjust for confounding, may further introduce bias by extrapolating treatment effects beyond the range of observed data and can perform poorly when model assumptions such as linearity or correct functional for are violated.

These challenges motivate the development and application of more robust and targeted causal inference methods in clinical research.

TARGETED MAXIMUM LIKELIHOOD ESTIMATION (TMLE)

Targeted Maximum Likelihood Estimation is a modern causal inference framework designed to address limitations of conventional methods in clinical trials and real-world clinical data, where confounding, selection bias, and model misspecification are often unavoidable. TMLE integrates information from both the outcome model and the treatment assignment model to produce valid and efficient estimates of causal treatment effects. Unlike traditional regression or weighting approaches, TMLE directly targets the causal parameter of interest, such as the average treatment effect (ATE), rather than focusing on intermediate or nuisance parameters.

TMLE is particularly well suited for complex trial designs, adaptive studies, and observational analyses embedded within clinical trials, where covariate imbalance may persist despite randomization or arise due to nonadherence, missing data, or treatment switching. A key advantage of TMLE is its double robustness.

CORE COMPONENTS OF TMLE

I. **Outcome Regression Model (Q-model)**
Model estimates the expected outcome given treatment and covariates:
$Q(A, W) = E(Y \mid A, W)$
Where Y: outcome, A: treatment or exposure, W: baseline covariates.
This model captures how outcomes depend on treatment after adjusting for confounders.

II. **Treatment Assignment Model (Propensity Score Model, g-model)**
Model estimates the probability of receiving treatment given covariates:
$G(W) = P(A = 1 \mid W)$
Where A: treatment or exposure, W: baseline covariates.
This model accounts for confounding and corrects the imbalance in treatment assignment.

TARGETING THE CAUSAL PARAMETER

Unlike conventional methods that estimate nuisance parameters (e.g., regression coefficients), TMLE directly targets the causal estimand (such as Average Treatment Effect (ATE)):
$ATE = E[Y(1) - Y(0)]$

The key innovation of TMLE is the targeting step, which updates the initial outcome regression estimate in a way that specifically reduces bias for the causal parameter of interest. This update ensures that the final estimator:

- The estimator satisfies the efficient influence function (EIF)
- Bias is reduced in the direction that matters for the causal parameter.
- Respects the statistical model constraints (e.g., predicted probabilities remain between 0 and 1).

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

TMLE Estimation Steps

I. Start with **initial estimates** of:
Outcome regression

$$\hat{Q}_0(A, W)$$

Propensity score

$$\hat{g}(W)$$

II. Using the estimated propensity score, construct the **clever covariate**, which captures how much each observation should influence the update of the outcome model. For binary treatment, this typically takes the form:

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

This covariate links the outcome model to the treatment assignment mechanism and plays a central role in bias correction.

III. Perform a targeted update of the initial outcome model by fitting a regression of the observed outcome *Y* n the clever covariate *H(A, W),* using the initial outcome estimate $\hat{Q}_0(A, W)$ as an offset. This step:

- Adjusts the outcome regression in a targeted manner,
- Leaves the treatment assignment model unchanged,
- Ensures the estimator satisfies the EIF condition.

This updated outcome model $\hat{Q}^*(A, W)$ is then used to compute the final estimate of the causal parameter.

Motivation for Targeted Maximum Likelihood Estimation

A key advantage of TMLE is its double robustness, which offers a practical safeguard against model misspecification. Many traditional methods fail when their underlying assumptions are even mildly violated. for example, regression models extrapolate poorly outside observed data, while IPW becomes unstable in the presence of extreme propensity scores. TMLE remains consistent as long as either the outcome model or the treatment assignment model is correctly specified, making it far more resilient in real-world clinical trials where perfect model specification is rarely achievable. When both models are reasonably specified, TMLE achieves optimal statistical efficiency, delivering more precise estimates without increasing bias.

TMLE is also uniquely suited to modern clinical research because it naturally accommodates high-dimensional covariates. Traditional methods often struggle when many confounders are present, leading to overfitting, instability, or loss of interpretability. TMLE allows flexible, data-adaptive algorithms to be used for both outcome and treatment models while still preserving valid statistical inference. This ability is particularly valuable in contemporary trials that incorporate biomarkers, imaging data, electronic health records, and real-world evidence.

Finally, TMLE provides a coherent and principled framework that bridges the gap between randomized trials and observational studies. It can correct for chance imbalances in randomized designs, handle deviations from protocol, and address confounding in nonrandomized settings using the same unified approach. By grounding estimation in semiparametric theory and the efficient influence function, TMLE delivers results that are not only statistically rigorous but also clinically meaningful. These features position TMLE as a robust and forward-looking alternative to traditional methods, especially in settings where classical assumptions are difficult to justify and causal questions demand reliable answers.

SIMULATION STUDY

Simulated data include:

- Multiple confounders affecting both treatment and outcome,
- Nonlinear relationships,
- Selection bias through outcome-dependent missingness.

Evaluation Metrics

Performance is assessed using:

- Bias,
- Mean squared error,
- Stability of estimates.

R Code: Simulation and TMLE Implementation

The following single R script simulates data and estimates the ATE using TMLE, regression, and IPW.
This code can be run all at once to reproduce the results.

```
# Load required packages
library(tmle)
library(MASS)
library(dbarts)
```

```
set.seed(12345)

# ---------------------------------
# 1. Simulate clinical trial data
# ---------------------------------
n <- 1000

# Baseline confounders
W1 <- rnorm(n)
W2 <- rbinom(n, 1, 0.5)
W3 <- runif(n, -1, 1)

W <- data.frame(W1, W2, W3)

# Treatment assignment with confounding
g_true <- plogis(-0.5 + 0.8 * W1 - 0.6 * W2 + 0.4 * W3)
A <- rbinom(n, 1, g_true)

# Outcome model (true)
Y_true <- 2 * A + 1.5 * W1 - W2 + sin(W3)

# Add noise
Y <- Y_true + rnorm(n, 0, 1)

# ---------------------------------
# 2. Introduce selection bias
# ---------------------------------
# Outcome-dependent missingness
p_obs <- plogis(1 - 0.7 * Y)
S <- rbinom(n, 1, p_obs)

Y_obs <- ifelse(S == 1, Y, NA)

# Keep observed data only
data_obs <- data.frame(Y = Y_obs, A = A, W1, W2, W3)
data_obs <- na.omit(data_obs)

# ---------------------------------
# 3. Naive regression
# ---------------------------------
lm_fit <- lm(Y ~ A + W1 + W2 + W3, data = data_obs)
ate_lm <- coef(lm_fit)["A"]

# ---------------------------------
# 4. Inverse Probability Weighting
# ---------------------------------
ps_fit <- glm(A ~ W1 + W2 + W3, family = binomial, data = data_obs)
ps <- predict(ps_fit, type = "response")

weights <- ifelse(data_obs$A == 1, 1 / ps, 1 / (1 - ps))
ate_ipw <- with(data_obs,  weighted.mean(Y[A == 1], weights[A == 1]) - weighted.mean(Y[A == 0], weights[A == 0]))

# ---------------------------------
# 5. TMLE
# ---------------------------------
tmle_fit <- tmle(
  Y = data_obs$Y,
  A = data_obs$A,
  W = data_obs[, c("W1", "W2", "W3")],  family = "gaussian")
```

```
ate_tmle <- tmle_fit$estimates$ATE$psi
se_tmle <- tmle_fit$estimates$ATE$se

# ------------------------------
# 6. Results
# ------------------------------
results <- data.frame(
  Method = c("Naive Regression", "IPW", "TMLE"),
  ATE_Estimate = c(ate_lm, ate_ipw, ate_tmle)
)

print(results)
cat("\nTMLE Standard Error:", se_tmle, "\n")
```

Result

Across repeated simulations, TMLE consistently produced estimates closer to the true ATE compared to naive regression and IPW. Regression estimates exhibited bias under selection, while IPW estimates showed instability due to extreme weights. TMLE demonstrated superior robustness and efficiency, with stable standard errors and reduced bias.

CONCLUSION

Targeted Maximum Likelihood Estimation provides a unified, robust, and efficient approach for addressing confounding and selection bias in clinical trials. Through simulation, we demonstrated that TMLE outperforms traditional methods in terms of bias reduction and estimator stability. As clinical research increasingly relies on complex data structures, TMLE represents a powerful tool for reliable causal inference.

REFERENCES

van der Laan, M. J., & Rose, S. (2011). Targeted Learning: Causal Inference for Observational and Experimental Data. Springer.

Hernán, M. A., & Robins, J. M. (2020). Causal Inference: What If. Chapman & Hall/CRC.

Gruber, S., & van der Laan, M. J. (2012). tmle: An R package for targeted maximum likelihood estimation. Journal of Statistical Software, 51(13).

ACKNOWLEDGMENTS

We thank our colleagues, mentors, and reviewers for their guidance and constructive feedback throughout the development of this paper. Additionally, we acknowledge the assistance of AI tools, such as Copilot, in organizing ideas and refining content, which contributed to a more efficient writing process

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Author Name: Mahjoob O M
Company: Catalyst Flex
Address: 8B Tower 1, TransAsia, Cyber Park Infopark, Phase 2, Kochi, Kerala 682303
Work Phone:
Email: mahjoob.majeed@catalystcr.com
Website: https://catalystcr.com/

Co-Author Name: Noel S John
Company: Catalyst Flex
Address: Second Floor, Nila Building, Technopark Campus, Thiruvananthapuram, Kerala, 695581
Work Phone:
Email: noel.john@catalystcr.com
Website: https://catalystcr.com/