

VARIABLE SELECTION IN RISK DIFFERENCE ESTIMATION

XUAN LI

ABSTRACT. This report serves as a qualifying paper for the UBC Statistics PhD program, providing a comprehensive summary of the paper “Variable Selection When Estimating Effects in External Target Populations” (Webster-Clark et al., 2024). The original paper shows that, to generalize risk difference estimates from a study population to a target population, one should include effect measure modifiers (EMMs) that differ between the populations; while adding non-EMMs differing between the populations might reduce precision, it does not amplify bias. Through simulations, we show that accounting for critical EMMs—both with and without interaction terms—consistently yields unbiased RD estimates, even when treatment assignment is not random. Moreover, we find that including interactions improves precision slightly but does not necessarily improve bias under the current experimental setup, suggesting potential areas for further methodological refinement.

CONTENTS

1. Introduction	1
1.1. Background and Motivation	1
1.2. Summary of the Original Paper	2
1.3. Empirical Simulation in the Paper	2
1.4. Simulation Results	3
1.5. Limitations	3
1.6. Exploring the Theoretical Approach	4
2. Potential Extensions	6
2.1. 1. Non-Random Treatment Assignment in the Target Population	6
2.2. 2. Assessing Performance When Key Interactions is Included	6
3. Exploration Through Simulation	6
3.1. Simulation Results	9
3.2. Conclusion	11
References	14

1. INTRODUCTION

1.1. Background and Motivation. Causal inference often requires estimating the treatment effects of study populations and generalizing them to target populations. One commonly used measure is the **risk difference (RD)**, which quantifies the difference in outcome probability between exposed and unexposed groups:

$$(1.1) \quad RD = P(Y = 1 \mid X = 1) - P(Y = 1 \mid X = 0).$$

In an ideal scenario, we can estimate RD using a study sample and apply it to a target population. However, differences in covariate distributions between study and target populations can lead to biased estimates.

A key challenge is selecting the appropriate covariates when modeling RD. An effect measure modifier (EMM) is a variable that changes the relationship between exposure and outcome (VanderWeele, 2010), in other words, it affects how exposure influences outcome. Non-effect measure modifiers (non-EMMs) do not modify the relationship between the exposure and the outcome, but can still affect the outcome independently. Understanding how different sets of variables influence model estimates in an external target population can enhance generalizability and improve the accuracy of effect estimation.

1.2. Summary of the Original Paper. The paper “*Variable Selection When Estimating Effects in External Target Populations*” primarily examines how covariate selection affects RD estimation when transporting causal effects from a study population to an external target population. The authors explore different modeling approaches, including **outcome modeling (OM)** and **inverse odds weighting (IOW)**, to assess their impact on bias and precision.

The authors examined various covariate sets and concluded that **effect measure modifiers (EMMs) differing across study and target populations** must be included—omitting them increases bias in transporting the RD to the target population. Meanwhile, **adding non-EMMs that differ** across populations inflates variance but does **not** amplify bias. These results challenge common intuitions from the internal validity setting and the discrepancy is explained by two main reasons from the authors:

- **Internal vs. External Validity Focus.** In **internal validity** (e.g., within a single study), one aims for conditional exchangeability between treated and untreated individuals, so adding strong outcome predictors (regardless of whether they modify the effect) can boost efficiency (Westreich et al., 2019). In **external validity**, the key is conditional exchangeability between *populations*, so the real analog of “outcome predictors” are the **EMMs**. Non-EMMs don’t mitigate external bias and merely increase variance.
- **Instrument-Like Variables vs. Non-EMMs.** In internal validity, instruments (variables related to the exposure) can amplify bias if unmeasured confounders are present (Myers et al., 2011). However, in external validity, the situation differs since the role of these variables differs. Here, non-EMMs that vary between study and target populations primarily reflect **sampling mechanisms** rather than confounding structures. As a result, we don’t observe bias amplification from non-EMMs in external validity as we do with instruments in internal validity.

1.3. Empirical Simulation in the Paper. The authors consider six covariates Z ’s in the original paper for simulation. The six covariates differ in two main aspects: (1) their distributions across the study and target populations, and (2) whether they affect the outcome Y and/or modify the effect of X on Y (i.e., whether they are effect modifiers).

The study focuses on binary treatment X and binary outcome Y . The true underlying process (DGP) of $P(Y|X, \mathbf{Z})$ with interaction terms of X and Z ’s is:

$$P(Y|X, \mathbf{Z}) = 0.1 + 0.1 \times X + 0.1 \times Z_2 + 0.1 \times Z_3 + 0.1 \times Z_5 + 0.1 \times Z_6 + 0.1 \times X \times Z_3 + 0.1 \times X \times Z_6,$$

where Z ’s represent other covariates besides treatment assignment X . In the paper’s simulation set-up, the treatment assignment X is random in both the target and the study data generation process. Note that Z_3, Z_6 **are EMMs**, as they interact with X in the DGP. Z_2, Z_5 **are non-EMMs**, affecting Y but not modifying the effect of X on Y . Z_1 and Z_4 are variables that neither affect Y nor affect how X influence Y . In the setup, Z_1, Z_2 and Z_3 are distributed in the same way across the study and target populations, while Z_4, Z_5 and Z_6 are distributed differently between the two populations.

The authors evaluate different linear model specifications for RD estimation under both **outcome modeling (OM)** and **inverse odds weighting (IOW)**. Below is a summary of key models used in their analysis in Table 1. Each of these models is applied in both the OM and IOW approaches. For the Outcome Modeling (OM) approach, linear models without interaction terms are fitted to stratified X in the study data and used to predict RD in the target population. For the inverse odds weighting (IOW) approach, linear weights are assigned to account for the different distributions of covariates between the study and target populations.

Model	Specification	Covariates Included
M_{crude}	Crude Model	X only
M_1	Non-EMM Z_1 only	X, Z_1
M_2	Non-EMM Z_2 only	X, Z_2
M_3	EMM Z_3 only	X, Z_3
\vdots	\vdots	\vdots
$M_{1,6}$	Non-EMM Z_1 + EMM Z_6	X, Z_1, Z_6
$M_{2,6}$	Non-EMM Z_2 + EMM Z_6	X, Z_2, Z_6
$M_{3,6}$	EMM Z_3 + EMM Z_6	X, Z_3, Z_6
\vdots	\vdots	\vdots

TABLE 1. Models Tested in the Paper

In the set-up, 10,000 data are generated in study population and 10,000 data are generated in target population. The authors conduct 20,000 simulations in total. Additionally, the authors also perform sensitivity analyzes in different study-target sizes and risk metrics.

1.4. Simulation Results. Empirically, the study finds that including non-EMMs does not amplify bias but reduces precision in both the OM and IOW approaches. The minimally sufficient adjustment set includes only Z_6 , the EMM that distributes differently between the study and target populations. Including non-EMMs identical between study and target has no effect, while including non-EMMs differing between populations besides the key EMM (such as including both Z_1 and Z_6) increases the empirical standard error (or standard deviation of the RD estimates across 20000 simulations) compared to only including Z_6 . When Z_6 is omitted, all estimates are biased ($RD = 0.170$), but adjusting for additional covariates does not amplify this bias compared to the crude model. Outcome modeling (OM) suffers less precision loss than inverse odds weighting (IOW). The sensitivity analyzes further confirm these findings.

In summary, the paper suggests that including only EMMs that differ between study and target is sufficient to obtain unbiased RD estimates. We should avoid adjusting for Non-EMMs that differ across populations.

1.5. Limitations. Although the authors provide valuable empirical results, their study has several limitations. First, all models they tested are in fact **misspecified** because none includes any interaction term $X \times Z_3$ or $X \times Z_6$, which are present in the true data-generating process (DGP). This means that all RD estimates come from **"incorrectly"** specified models, raising the question of how meaningful the comparisons between these models truly are and whether similar patterns would emerge if the models correctly accounted for interaction effects. Second, the simulation setup assumes that treatment X

is randomly assigned in the target population, an assumption that often does not hold in real-world applications. In practice, exposure is typically influenced by the underlying covariates, which introduces potential selection bias that the study does not account for. Third, the study lacks a theoretical derivation of the bias and variance properties of RD estimators. Although empirical findings suggest that including non-EMMs increases variance without introducing bias, there is no formal mathematical justification supporting this conclusion.

Other limitations include the use of a **linear probability model** instead of the more commonly used logistic model for binary outcomes, which may lead to less accurate probability estimates, especially when outcome probabilities are near 0 or 1. Additionally, the study does not incorporate confounding effects, assuming no unmeasured common causes of both X and Y . In many practical settings, unmeasured confounding can bias causal effect estimates, making it an important factor to consider in generalizing RD estimates. Given these limitations, our report aims to further investigate several of these issues and assess their implications for RD estimation in external validity settings.

1.6. Exploring the Theoretical Approach. While the main findings of the paper rely on empirical simulations, we try to explore a simplified theoretical framework to understand why including non-EMMs does not increase bias but does decrease precision. Consider a simplified version of the DGP for Y given exposure X and covariates Z_1, Z_2 used in the original paper:

$$(1.2) \quad P(Y = 1 \mid X, Z_1, Z_2) = \beta_0 + \beta_1 \times X + \beta_2 \times Z_1 + \beta_3 \times X \times Z_2$$

where Z_1 is a non-EMM, and Z_2 is an EMM. In this case, the true RD, defined as $RD = P(Y = 1 \mid X = 1) - P(Y = 1 \mid X = 0)$ simplifies to:

$$(1.3) \quad RD = \beta_1 + \beta_3 Z_2.$$

However, as done in the simulation from the original paper, we do not fit the true model; instead, we fit **misspecified models** that omit interaction terms. We explore the bias and variance of RD under three model specifications.

Model 1: Suppose we fit a linear model that includes X and Z_1 , but omits Z_2 and its interaction with X :

$$(1.4) \quad \hat{P}(Y = 1 \mid X, Z_1) = \hat{\beta}_0 + \hat{\beta}_1 X + \hat{\beta}_2 Z_1.$$

The **estimated RD** from this model is:

$$(1.5) \quad \hat{RD} = \hat{\beta}_1,$$

The **bias** of estimated RD is therefore:

$$(1.6) \quad \text{Bias}(\hat{RD}) = (\beta_1 - \hat{\beta}_1) + \beta_3 E[Z_2].$$

The term $\beta_1 - \hat{\beta}_1$ arises from **estimation error** and the term $\beta_3 E[Z_2]$ is the bias introduced by omitting the interaction term $X \times Z_2$. The variance of estimated RD in this case is given by:

$$(1.7) \quad \text{Var}(\hat{RD}) = [(X^T X)^{-1}]_{11} \sigma^2.$$

where σ^2 is the residual variance and $[(X^T X)^{-1}]_{11}$ is the entry in the inverse matrix $(X^T X)^{-1}$ corresponding to β_1

Model 2: Now consider an alternative model that includes X and Z_2 , but omits Z_1 :

$$(1.8) \quad \hat{P}(Y = 1 \mid X, Z_2) = \hat{\beta}_0^* + \hat{\beta}_1^* X + \hat{\beta}_3^* Z_2.$$

Here, the **estimated RD** is:

$$(1.9) \quad \hat{RD}^* = \hat{\beta}_1^*.$$

The **bias** of estimated RD in this model is:

$$(1.10) \quad \text{Bias}(\hat{RD}^*) = (\beta_1 - \hat{\beta}_1^*) + \beta_3 E[Z_2].$$

The term $\beta_1 - \hat{\beta}_1^*$ is again due to **estimation error** and the term β_3 is due to omitting the interaction term $X \times Z_2$. The **variance** of estimated RD in this case is:

$$(1.11) \quad \text{Var}(\hat{RD}^*) = [(X^T X)^{-1}]_{11}^* \sigma^2.$$

Model 3: Now consider another model that includes X , Z_1 and Z_2 :

$$(1.12) \quad \hat{P}(Y = 1 | X, Z_2) = \hat{\beta}_0^\# + \hat{\beta}_1^\# X + \hat{\beta}_2^\# Z_1 + \hat{\beta}_3^\# Z_2.$$

Here, the **estimated RD** is:

$$(1.13) \quad \hat{RD}^\# = \hat{\beta}_1^\#.$$

The **bias** of estimated RD in this model is:

$$(1.14) \quad \text{Bias}(\hat{RD}^\#) = (\beta_1 - \hat{\beta}_1^\#) + \beta_3 E[Z_2].$$

The term $\beta_1 - \hat{\beta}_1^\#$ is again due to **estimation error** and the term β_3 is due to omitting the interaction term $X \times Z_2$. The **variance** estimated of RD in this case is:

$$(1.15) \quad \text{Var}(\hat{RD}^\#) = [(X^T X)^{-1}]_{11}^\# \sigma^2.$$

From the theoretical expressions above, we can make the following key observations:

- **Bias Analysis:** The difference in bias across three estimated RD's depends on how accurately $\hat{\beta}_1$ is estimated. Since all three models are misspecified, it is difficult to determine which provides a more accurate estimate of β_1 . However, the bias does not introduce any additional terms due to including a non-EMM (Z_1).
- **Variance Analysis:** Including additional non-EMMs increases the dimensionality of the matrix X , and if these covariates are correlated with the treatment variable, they can inflate the elements of $[X^T X]^{-1}$. This results in a larger variance for the estimated $\hat{\beta}_1$, decreasing precision.

Thus, while the theoretical approach does not allow us to compare bias between models in a clear way, it supports the empirical finding that including non-EMMs decreases precision. Since theoretical derivations cannot easily compare which misspecified model yields a better estimate of β_1 , simulation experiments remain crucial to evaluate bias performance in practice.

It is important to notice that the above derivations are intended as a heuristic illustration rather than a formal proof. We automatically assume certain conditions, such as, the **same data-generating process** applies in both the study and target populations (so that comparing bias of transported estimates across them is valid), the **covariates are not highly collinear** (so the inverse of $X^T X$ is stable), and a **linear probability model** is an acceptable approximation. In reality, these assumptions may not hold, which is why simulation experiments remain crucial to confirm the bias and variance performance in practical scenarios.

2. POTENTIAL EXTENSIONS

The original study provides valuable insights into variable selection for external generalization but leaves several key questions open, such as the assumption on random exposure assignment in the target population, and the consequences of misspecification due to omitted interaction terms in the model. We propose two extensions as potential research directions: (1) studying non-random exposure assignment and its impact on RD estimation, and (2) assessing how the omission of interaction terms affects bias and variance in different modeling approaches.

2.1. 1. Non-Random Treatment Assignment in the Target Population. The original study assumes that the binary treatment assignment X follows a Bernoulli(0.5) distribution in both the study and target populations, meaning that individuals in both settings have an equal probability of receiving the treatment. However, in most real-world applications, treatment assignment is not random but rather influenced by underlying characteristics, such as socioeconomic status, regional healthcare access, or personal health history in a target population. For example, in the real world, wealthier individuals may have better access to treatment than lower-income individuals if the treatment is very costly.

Why This Matters: When treatment assignment mechanisms differ between the study and target populations, standard IOW or OM methods may fail to account for confounding introduced by non-random treatment assignment. If X is influenced by a non-EMM in the target population but not in the study, or vice versa, then bias in RD estimation may persist. Investigating how OM, IOW, or new weighting approaches, performs under non-random treatment assignment and assessing whether additional covariate adjustment can mitigate potential bias is meaningful for ensuring robust RD estimation in more realistic settings.

2.2. 2. Assessing Performance When Key Interactions is Included. The models tested in the paper do not account for interaction terms between X and effect modifiers (EMMs) such as Z_3 and Z_6 , even though the true data-generating process includes such interactions. This means that all fitted models in the study are misspecified with respect to the true underlying causal relationship. For example, omitting the interaction $X \times Z_6$ implicitly assumes that Z_6 affects Y equally for both exposed ($X = 1$) and unexposed ($X = 0$) individuals, which may not be valid.

Why This Matters: Including interaction terms is essential when estimating causal effects in the presence of effect modification. Ignoring these terms can lead to biased RD estimates. Investigating how interaction effects influence models can provide insights into the robustness of these methods when used in practice. Additionally, evaluating whether alternative methods, such as doubly robust estimators, can correct for these biases would be valuable.

3. EXPLORATION THROUGH SIMULATION

We conduct three simulations to assess the performance of outcome modeling (OM) under different data-generating mechanisms while simplifying the original study by omitting inverse odds weighting (IOW). The setup for covariates Z_1 to Z_6 and the outcome model $P(Y|X, \mathbf{Z})$ follows the same structure as the original paper, ensuring consistency in estimating the risk difference (RD). Additionally, We introduce Z_7 and Z_8 to investigate how non-random treatment assignment impacts generalizability when underlying covariates have different distributions across study and target populations.

In the first simulation, X is assigned randomly in both the study and target populations like in the original paper. For the other two simulations, we allow X to be assigned based on underlying covariates Z_7 and Z_8 . The probability of treatment assignment in the second simulation $P(X)$ depends on Z_7 and Z_8 , with Z_7 following different binomial distributions in the study and target populations while Z_8 remained identically distributed. The third simulation extended this by modeling Z_7 and Z_8 as continuous variables drawn from beta distributions. Figures 1 to 3 and Tables 2 to 4 illustrate the distribution of $X = 1$ and $X = 0$ in the study and target populations for example simulations of the three different scenarios. In Figure 1, X is randomly assigned, resulting in balanced distributions across both populations, this is further confirmed by Table 2. In contrast, as shown in Figures 2 and 3, and Tables 3 and 4, when $P(X)$ is influenced by Z_7 and Z_8 , the distribution of X leads to an imbalance between the study and target populations. These figures serve as illustrative examples from a larger simulation study, where we conducted 500 simulations for each three simulation scenarios to estimate the mean, standard deviation and mean squared error of RD estimates using different model specifications.

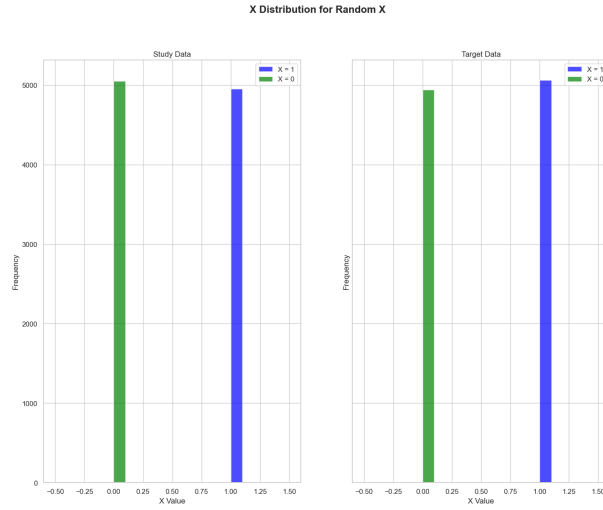


FIGURE 1. The histogram shows the distribution of $X = 1$ and $X = 0$ in the study and target data in the first simulation scenario, where X is randomly assigned. The left panel represents the study data, while the right panel represents the target data.

Left-aligned	Centered	Right-aligned
Dataset	Mean	Std
Study Data	0.496	0.25
target Data	0.506	0.25

TABLE 2. The table represents the mean and standard deviation of X in the study and target data when X is randomly assigned in both datasets. The values are rounded to 3 decimal places.

Across all simulations, Z_3 and Z_6 are treated as effect modifiers (EMMs), but only Z_6 differs in distribution between the study and target populations. We estimate RD using three modeling approaches: (1) outcome models without interaction terms, (2) models including only the $X \times Z_6$ interaction, and (3) models including interaction terms $X \times Z_3$ and $X \times Z_6$. Linear terms for covariates involved in interactions are always included in models containing interaction terms to ensure correct model specification and proper interpretation of the interaction effects. For example, if the model only interacts with Z_6 ,

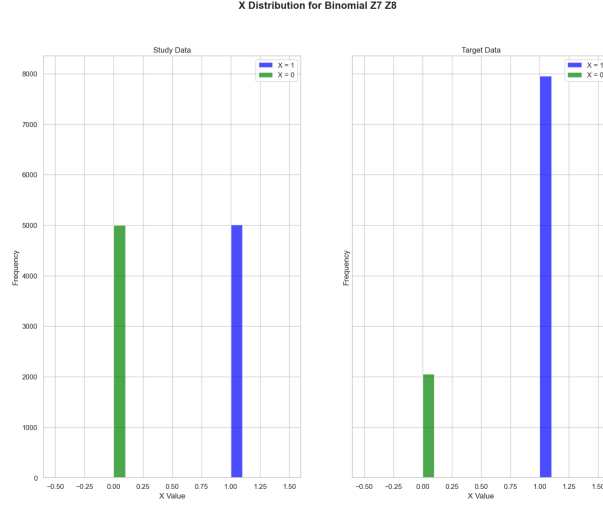


FIGURE 2. The histogram shows the distribution of $X = 1$ and $X = 0$ in the study and target data for the second simulation scenario, where $P(X)$ in the target data is determined by Z_7 and Z_8 , which are binomially distributed. The left panel represents the study data, while the right panel represents the target data.

Left-aligned	Centered	Right-aligned
Dataset	Mean	Std
Study Data	0.501	0.25
target Data	0.795	0.163

TABLE 3. The table represents the mean and standard deviation of X in the study and target data when X is not randomly assigned in the target data. The values are rounded to 3 decimal places. In this case, $P(X)$ in the target data depends on Z_7 and Z_8 . Z_7 and Z_8 follow binomial distributions.

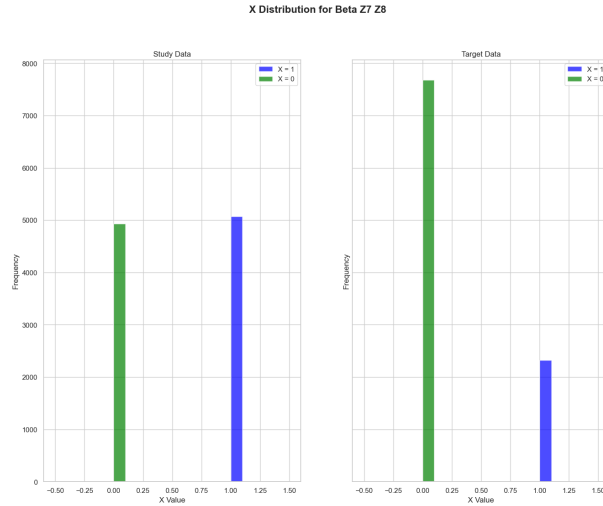


FIGURE 3. Distribution of $X = 1$ and $X = 0$ in the study and target data for the third simulation scenario, where $P(X)$ in the target data is determined by Z_7 and Z_8 , which follows beta distribution. The left panel represents the study data, while the right panel represents the target data.

there is always a linear Z_6 term. If the model interacts with Z_6 and Z_3 , there are always a linear Z_6 term and a linear Z_3 term.

Left-aligned	Centered	Right-aligned
Dataset	Mean	Std
Study Data	0.507	0.25
target Data	0.232	0.178

TABLE 4. The table represents the mean and standard deviation of treatment assignment X in study data and target data when X is not randomly assigned in the target data. The values are rounded to 3 decimal places. In this case, $P(X)$ in the target data depends on Z_7 and Z_8 . Z_7 and Z_8 follows beta distributions.

In all three approaches, we use different combinations of linear terms involving the non-EMMs (i.e., Z_1, Z_2, \dots, Z_8) to testify how those variable selections would affect estimating RD. We followed the original paper in using stratified models for X (separate models for treated and untreated groups) when no interaction terms are involved; however, when interaction terms are included, we instead fit a single joint model that included X , selected covariates, and interaction terms.

3.1. Simulation Results. To assess the performance of different modeling strategies, we present:

- **Three tables** (Figures 4–6) summarizing the mean, standard deviation (Std), and mean squared error (MSE) of RD estimates across 500 simulations. These results compare models with and without interaction terms under three different data-generating processes (DGPs) for X and Z_7, Z_8 .
- **Two sets of line plots** (Figures 7 and 8), visualizing how MSE and Std vary across models. These plots contrast linear models (without interaction terms) that include the crucial EMM Z_6 against models incorporating interaction terms.

We aim to explore on three naturally raised questions: How does **random vs. non-random treatment assignment** affect RD estimates? Do **interaction terms** improve or worsen the estimate in terms of MSE and Std of the estimates? What is the **best modeling strategy** based on the results?

Impact of Random vs. Non-Random Treatment Assignment. Our results suggest that whether treatment assignment in the target population is random or not does not make a major difference in RD estimation, as long as the key effect measure modifier (EMM) Z_6 is included in the model. When Z_6 is omitted, we see a slight improvement in MSE under random exposure assignment. However, once Z_6 is included—either as a linear term or as an interaction term with X —this advantage disappears, though the standard deviation remains slightly lower when X is randomly assigned. The results are presented in Figures 4 to 6.

One possible explanation is that when the correct covariates are included, the model already accounts for systematic differences in exposure assignment between the study and target populations, making randomization in treatment assignment less important for RD estimation.

Overall, while random exposure assignment may provide a small advantage in standard deviation when the main EMM is included, its overall impact on RD estimation appears minimal. This suggests that careful covariate selection might be more important than the specific exposure assignment mechanism when generalizing causal effects.

Effect of Interaction Terms on Estimation Stability. We find that including interaction terms for EMMs slightly reduces variance compared to including an EMM as only a linear term. A possible explanation is that modeling a strong effect modifier as

Models with no Interaction term									
	Random X			Non-Random X (Binomial Z7 Z8)			Non-Random X (Beta Z7 Z8)		
	mean	std	mse	mean	std	mse	mean	std	mse
rd_intercept	0.1706	0.0092	0.0036	0.1704	0.0089	0.0036	0.1695	0.0091	0.0037
rd_Z1	0.1706	0.0092	0.0036	0.1704	0.0089	0.0036	0.1695	0.0091	0.0037
rd_Z2	0.1705	0.0091	0.0036	0.1704	0.0089	0.0036	0.1695	0.0090	0.0037
rd_Z3	0.1705	0.0091	0.0036	0.1704	0.0088	0.0036	0.1695	0.0090	0.0037
rd_Z4	0.1714	0.0169	0.0037	0.1707	0.0158	0.0038	0.1696	0.0165	0.0039
rd_Z5	0.1697	0.0173	0.0039	0.1709	0.0172	0.0038	0.1696	0.0169	0.0039
rd_Z6	0.2301	0.0169	0.0003	0.2301	0.0173	0.0003	0.2297	0.0173	0.0003
rd_Z7	0.1694	0.0168	0.0039	0.1710	0.0163	0.0037	0.1695	0.0091	0.0037
rd_Z8	0.1706	0.0092	0.0036	0.1704	0.0089	0.0036	0.1696	0.0092	0.0037
rd_Z1Z6	0.2300	0.0169	0.0003	0.2301	0.0172	0.0003	0.2297	0.0174	0.0003
rd_Z2Z6	0.2300	0.0165	0.0003	0.2299	0.0172	0.0003	0.2295	0.0172	0.0003
rd_Z3Z6	0.2303	0.0167	0.0003	0.2301	0.0170	0.0003	0.2296	0.0171	0.0003
rd_Z4Z6	0.2309	0.0220	0.0005	0.2304	0.0218	0.0005	0.2298	0.0224	0.0005
rd_Z5Z6	0.2291	0.0218	0.0005	0.2307	0.0233	0.0005	0.2300	0.0222	0.0005
rd_Z7Z6	0.2291	0.0217	0.0005	0.2307	0.0222	0.0005	0.2297	0.0173	0.0003
rd_Z8Z6	0.2300	0.0169	0.0003	0.2301	0.0173	0.0003	0.2297	0.0173	0.0003
true_rd	0.2298	0.0024	0.0000	0.2298	0.0026	0.0000	0.2298	0.0030	0.0000

FIGURE 4. Results for models without interaction terms, comparing random and non-random treatment assignment across three simulation settings introduced earlier. The first three columns correspond to Simulation 1, the middle three to Simulation 2, and the right three to Simulation 3. Notation Pattern: *rd_intercept* represents a model with only the intercept and X , *rd_Z1* includes the intercept, X , and Z_1 , and *rd_Z1Z6* includes the intercept, X , Z_1 , and Z_6 . *true_rd* is the RD calculated from the target population data. Lighter colors indicate lower bias and variance, while darker colors indicate higher values, reflecting model stability and performance.

Models with Interaction term $X \times Z_6$									
	Random X			Non-Random X (Binomial Z7 Z8)			Non-Random X (Beta Z7 Z8)		
	mean	std	mse	mean	std	mse	mean	std	mse
rd_intercept	0.2301	0.0169	0.0003	0.2301	0.0173	0.0003	0.2297	0.0173	0.0003
rd_Z1	0.2301	0.0169	0.0003	0.2301	0.0172	0.0003	0.2297	0.0173	0.0003
rd_Z2	0.2300	0.0165	0.0003	0.2299	0.0173	0.0003	0.2295	0.0172	0.0003
rd_Z3	0.2278	0.0167	0.0003	0.2275	0.0172	0.0003	0.2271	0.0172	0.0003
rd_Z4	0.2300	0.0169	0.0003	0.2301	0.0172	0.0003	0.2297	0.0173	0.0003
rd_Z5	0.2300	0.0168	0.0003	0.2301	0.0172	0.0003	0.2297	0.0172	0.0003
rd_Z7	0.2301	0.0169	0.0003	0.2301	0.0172	0.0003	0.2297	0.0173	0.0003
rd_Z8	0.2301	0.0169	0.0003	0.2301	0.0173	0.0003	0.2297	0.0173	0.0003
true_rd	0.2298	0.0024	0.0000	0.2298	0.0026	0.0000	0.2298	0.0030	0.0000

FIGURE 5. Results for models including the interaction term $X \times Z_6$, comparing random and non-random treatment assignment across three simulation settings introduced earlier. The first three columns correspond to Simulation 1, the middle three to Simulation 2, and the right three to Simulation 3. Notation Pattern: *rd_intercept* represents a model including the interaction term $X \times Z_6$, the linear term Z_6 , the intercept, and X . *rd_Z1* includes the same components as *rd_intercept*, plus the linear term Z_1 . *true_rd* is the RD calculated from the target population data. Lighter colors indicate lower bias and variance, while darker colors indicate higher values, reflecting model stability and performance.

an interaction term can help reduce residual (unexplained) variation, which **offsets** the variance cost of estimating an additional parameter, leading to more stable estimates.

However, the overall improvement in MSE is minimal, as models that include the crucial EMM as a linear term already exhibit low MSE. This suggests that, in the current simulation setting, omitting interaction effects does not induce substantial bias in RD estimation. Nonetheless, we observe that when interaction terms are omitted but non-EMMs (Z_4 , Z_5 , and Z_7) are included, both MSE and Std increase noticeably. Since these variables differ between the study and target populations, their inclusion without modeling interaction effects may introduce instability in RD estimation. Results are presented in Figures 7 and 8.

Models with Interaction terms $X \times Z_3$ and $X \times Z_6$									
	Random X			Non-Random X (Binomial Z7 Z8)			Non-Random X (Beta Z7 Z8)		
	mean	std	mse	mean	std	mse	mean	std	mse
rd_intercept	0.2303	0.0167	0.0003	0.2301	0.0170	0.0003	0.2296	0.0171	0.0003
rd_Z1	0.2303	0.0167	0.0003	0.2301	0.0170	0.0003	0.2296	0.0171	0.0003
rd_Z2	0.2302	0.0164	0.0003	0.2299	0.0170	0.0003	0.2294	0.0169	0.0003
rd_Z4	0.2303	0.0167	0.0003	0.2301	0.0170	0.0003	0.2296	0.0171	0.0003
rd_Z5	0.2303	0.0167	0.0003	0.2301	0.0169	0.0003	0.2296	0.0169	0.0003
rd_Z7	0.2303	0.0168	0.0003	0.2301	0.0170	0.0003	0.2296	0.0171	0.0003
rd_Z8	0.2303	0.0168	0.0003	0.2301	0.0170	0.0003	0.2296	0.0171	0.0003
true_rd	0.2298	0.0024	0.0000	0.2298	0.0026	0.0000	0.2298	0.0030	0.0000

FIGURE 6. Results for models including the interaction terms $X \times Z_6$ and $X \times Z_3$, comparing random and non-random treatment assignment across three simulation settings introduced earlier. The first three columns correspond to Simulation 1, the middle three to Simulation 2, and the right three to Simulation 3. Notation Pattern: *rd_intercept* represents a model including the interaction terms $X \times Z_6$ and $X \times Z_3$, the linear terms Z_6 and Z_3 , the intercept, and X . *rd_Z1* includes the same components as *rd_intercept*, plus the linear term Z_1 . *true_rd* is the RD calculated from the target population data. Lighter colors indicate lower bias and variance, while darker colors indicate higher values, reflecting model stability and performance.

Comparing models that include only $X \times Z_6$ versus those that include both $X \times Z_3$ and $X \times Z_6$, we find no meaningful difference in MSE. However, models incorporating both interaction terms exhibit lower Std. This suggests that while adding $X \times Z_3$ does not substantially alter the expected risk difference, it may help stabilize estimation by better capturing treatment effect heterogeneity.

Best modeling strategy. Based on our results, it is recommended that:

- (1) Always include the key EMM Z_6 to prevent bias.
- (2) Including EMMs (even if they do not differ across populations, like Z_3) **reduces variance** in the estimates, leading to **more stable and precise RD estimates**.
- (3) **Preferably** include EMMs as interaction terms when there is theoretical justification or empirical evidence supporting their role in the true data-generating process.
- (4) Avoid including non-EMMs that differ across populations unless necessary, as they may increase variance without reducing bias.
- (5) If non-EMMs must be included, ensure that interactions involving key EMMs are also modeled to mitigate variance inflation.

This strategy aims to balance bias reduction and variance stability, leading to the most reliable RD estimates in external validity settings.

3.2. Conclusion. Our findings align with the original paper’s conclusion that essential effect measure modifiers (EMMs)—particularly those differing between study and target populations—need to be included for unbiased RD estimate and including non-EMMs differ between populations decreases precision but does not amplify bias. While adding interaction terms does not substantially improve risk difference (RD) estimation in our simulations, this may be due to the simplicity of our experimental setup. A more realistic scenario could involve varying the magnitude of interactions or incorporating more complex structures—such as higher-order or nonlinear effects—to test the robustness of variable selection strategies.

Additionally, future studies should explore more complex modeling settings beyond interaction terms. Introducing selection bias, unmeasured confounding, or nonlinear dependencies could better reflect real-world transportability challenges. Since our results

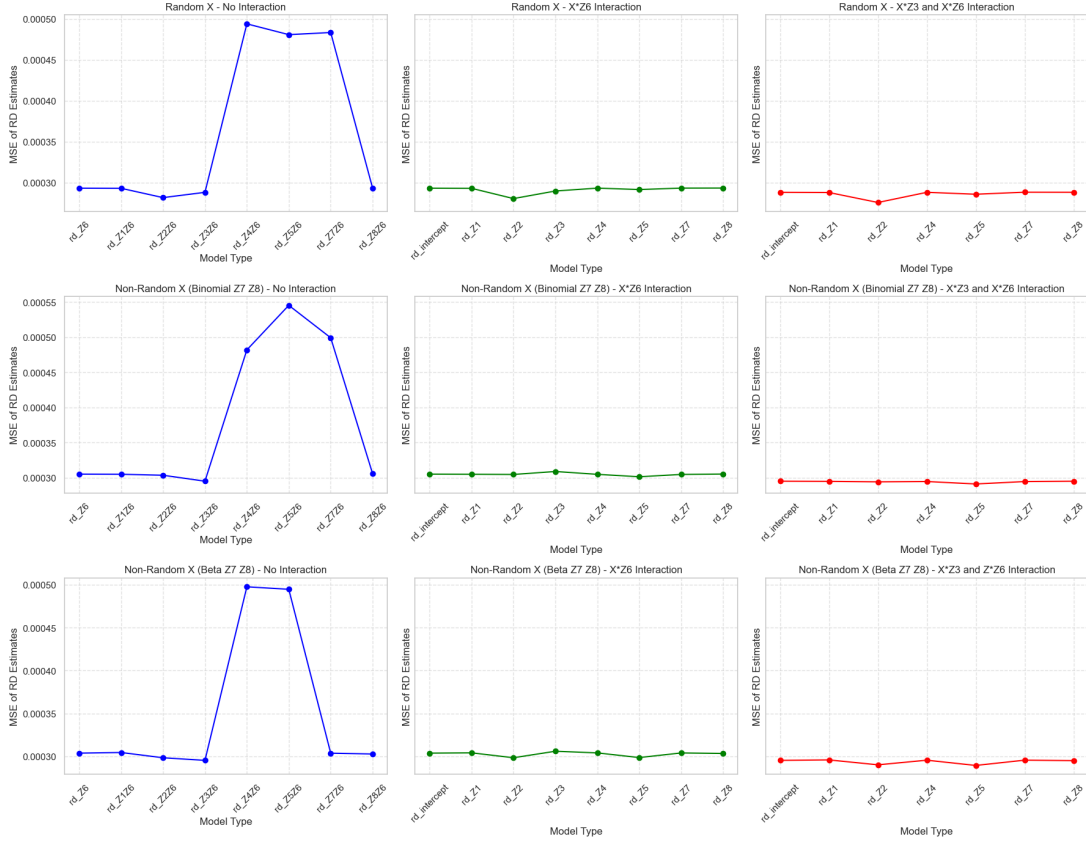


FIGURE 7. Line plots of Mean Squared Error (MSE) in the RD estimates under different model specifications. The first column shows models without interaction terms, the second column includes models with the interaction term $X \times Z_6$, and the third column incorporates both $X \times Z_3$ and $X \times Z_6$. The first row represents scenarios where X is randomly assigned in the target population, while the second and third rows correspond to non-random treatment assignment with different distributions of Z_7 and Z_8 influencing X .

indicate minimal gains from including interactions under the current setup, further research should assess whether a more sophisticated experimental design would show more distinct results.

Importantly, **identifying EMMs** in real-world data is challenging. As the original paper also emphasizes, prioritizing variables strongly associated with the outcome is often a reasonable approach, even if it comes at the cost of increased variance. This highlights an inherent bias-variance trade-off in selecting adjustment sets for external validity.

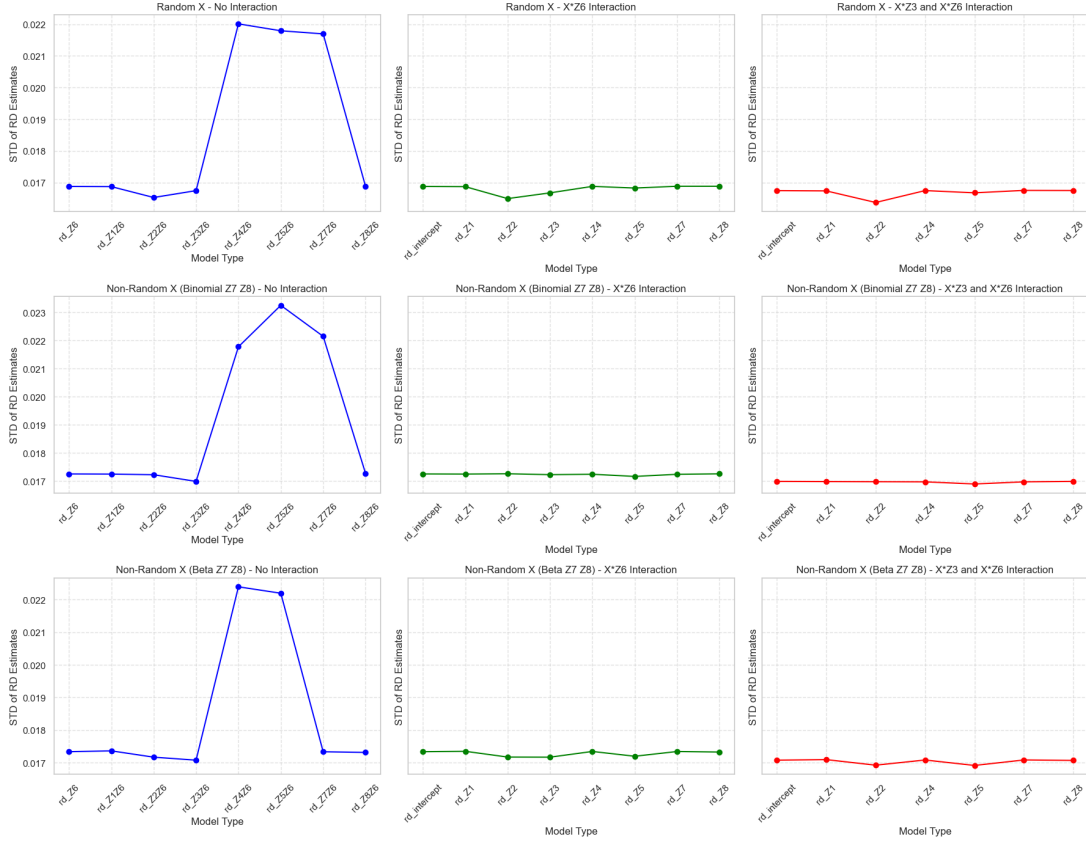


FIGURE 8. Line plots of Standard Deviation (Std) in the RD estimates under different model specifications. The first column shows models without interaction terms, the second column includes models with the interaction term $X \times Z_6$, and the third column incorporates both $X \times Z_3$ and $X \times Z_6$. The first row represents scenarios where X is randomly assigned in the target population, while the second and third rows correspond to non-random treatment assignment with different distributions of Z_7 and Z_8 influencing X .

REFERENCES

- Myers, J. A., Rassen, J. A., Gagne, J. J., Huybrechts, K. F., Schneeweiss, S., Rothman, K. J., Joffe, M. M., & Glynn, R. J. (2011). Effects of adjusting for instrumental variables on bias and precision of effect estimates. *Am J Epidemiol*, 174(11), 1213–1222.
- VanderWeele, T. J. (2010). On the distinction between interaction and effect modification. *Epidemiology*, 21(1), 162.
- Webster-Clark, M., Ross, R. K., Keil, A. P., & Platt, R. W. (2024). Variable selection when estimating effects in external target populations. *American Journal of Epidemiology*, 193, Issue 8.
- Westreich, D., Edwards, J. K., Lesko, C. R., Cole, S. R., & Stuart, E. A. (2019). Target validity and the hierarchy of study designs. *Am J Epidemiol*, 188(2), 438–443.