*Article*

# Treatment Benefit and Treatment Harm Rates with Nonignorable Missing Covariate, Endpoint, or Treatment

**Yi He, Linzhi Zheng and Peng Luo ***

School of Mathematics and Statistics, Shenzhen University, Shenzhen 518061, China;
2070205030@email.szu.edu.cn (L.Z.)
* Correspondence: luopeng@szu.edu.cn

**Abstract:** The average treatment effect is an important concept in causal inference. However, it fails to capture variation in response to treatment due to heterogeneity at many levels among patients in the target population. To study the heterogeneity in the treatment effect, researchers proposed the concepts of treatment benefit rate (TBR) and treatment harm rate (THR). Howerver, in practice, missing data often occurs in treatment, endpoints, and covariates. In these cases, the conditions given by them are not enough to identify treatment benefit rate. In this article, we address the problem of identifying the treatment benefit rate and treatment harm rate when treatment or endpoints or covariates are missing. Different types of missing data mechanisms are assumed, including several situations of nonignorable missingness. We prove that the treatment benefit rate and treatment harm rate are identifiable under very mild conditions, and then construct estimators based on methods of the EM algorithm. The performance of the proposed inference procedure is evaluated via simulation studies. Lastly, we illustrate our method by real data sets.

**Keywords:** nonignorable missing; treatment harm rates; treatment benefit rates; causal inference; nonparametric estimation

**MSC:** 62G05

## 1. Introduction

The average treatment effect is widely used in the measurement of causal inference [1]; however, it is not the only measure. For example, in a typical randomized Phase III clinical trial, there are patients who benefit from a negative trial and patients who do not benefit from a positive trial. In this situation, the average treatment effect can not explain the casual effect completely, since it ignores the heterogeneous responses to the treatment in the target population. Some researchers made additional assumptions to address this heterogeneity. One of the main assumptions is "Monotony" [2], which assumes that the treatment effect for each individual will be no worse than the control effect. There are many scientific and empirical reasons to doubt this assumption. Ref. [3] proposed several explanations for the fact that some people respond to an inactive control group, but do not respond to an experimental treatment, and note that for some people, a placebo has been shown to be superior to active treatment.

For this reason, we focus on the measurement of treatment benefit rate (TBR) and treatment harm rate (THR) in our paper. Ref. [4] tried to identify TBR and THR by making the additional assumption that the two potential outcomes were independent, conditional on observed covariates. Ref. [5] estimated the TBR and THR assuming the existence of at least three covariates, which are mutually independent. Ref. [6] proposed a Bayesian-tree-based latent variable model to seek subpopulations with distinct TBR. Under the assumption that the potential outcomes are independent conditional on the observed covariates and an unmeasured latent variable, ref. [7] showed the identification of the TBR and THR in non-separable (generalized) linear mixed models for both continuous and

binary outcomes. In our article, we follow the assumption in [4] to make the TBR and THR identifiable.

However, although our experiment is based on the assumption of a randomized experiment, we still need to face a problem in the process of identifying TBR and THR, that is, there may be missing data in the pretreatment covariate, potential endpoint, or treatment assignment. When one of the variables has missing data, TBR and THR cannot be identified; we can only get the upper and lower bounds for parameters of interest, rather than point estimates, and the upper and lower bounds are too wide to use. When the missing data mechanism is ignorable, we can ignore observations with missing data and identify TBR and THR directly. In many cases, however, missing data is not ignorable, that is, the processing of missing data depends on some possibly missing variables. In this case, TBR and THR cannot be identified without other assumptions. For example, in a randomized clinical trial [8], the covariate is obtained from electrophysiological stimulation (EPS) testing. Because the EPS testing is invasive and not a prerequisite for enrollment in the study, 79.3% of patients in the implantable cardiac defibrillator arm have EPS records, whereas only 2.4% of patients in the control arm have EPS records. Therefore, the missing data problem for the covariate is very severe and nonignorable. Another example is the Awakening and Breathing Controlled trial [9]. In this trial, because of the possibility of the patients' death, there are nonignorable missing data in the cognitive score at 3 months and 12 months. In a total of 187 patients, there were 111 missing values for the cognitive score at 3 months and 136 missing values for the cognitive score at 12 months, and they are nonignorable. Because of the missing data, we can not identify the TBR and THR directly.

There are many examples in the literature where the problem of missing data in causal inference has been studied. Refs. [8,10] used sensitivity analysis in the nonignorable missing covariates problem. Refs. [11–13] studied the identification problem when the missingness of the outcomes was nonignorable. Refs. [14–16] discussed the identifiability of causal effects when a key covariate is missing due to death. Refs. [17,18] also discussed nonignorable missing covariates problems in survival analysis and regression models. In our paper, we deal with the case that one of the treatments, covariate, and the endpoint have missing data. We will give some basic assumptions and special conditions for pretreatment covariates, potential endpoints, and treatment, under which we can identify the TBR and THR. These assumptions and conditions have certain wide applicability.

The rest of this article is as follows. In Section 2, we introduce the notation and assumption used throughout this article. In Section 3, we introduce several missing mechanisms of covariate, endpoint, and treatment. In Section 4, we discuss the identifiability of TBR and THR under these missing mechanisms. In Section 5, we estimate TBR and THR using the EM algorithm in simulation studies when they can be identified. In Section 6, we analyze datasets from clinical trials by our methods. Lastly, we put the proofs of theorems in the Appendix A.

## 2. Notation and Assumption

Let $Z$ denote the treatment assignment. $Z = 1$ means treatment, and $Z = 0$ means control. We assume that there is only one covariate, and let $X$ denote the pretreatment covariate with $K$ categories($K \geq 2$). Suppose the $K$ levels of $X$ are $x_1, x_2, \ldots x_K$. Let $Y$ donate endpoints, and suppose $Y$ is binary. $Y = 1$ means that the treatment or control works. $Y = 0$ means that the treatment or control does not work. We assume $Y(0)$ as the potential endpoint under control and $Y(1)$ as the potential endpoint under intervention. Then, the observed endpoint $Y$ can be written as $Y = ZY(1) + (1 - Z)Y(0)$. In our article, we assume that one of $X, Y, Z$ is missing. Let $R_X$ denote the missing data indicator for $X$ and $R_X(z)$ denote the potential missing data indicator for $X$, $R_Y$ denotes the missing indicator for $Y$ and $R_Y(z)$ denotes the potential missing data indicator for $Y$, $R_Z$ denotes the missing indicator for $Z$. Because $Z$ is the treatment, we do not write the potential variable of $R_Z$. We can only observed one of the pairs $\{Y(1), R_X(1), R_Y(1)\}$ and $\{Y(0), R_X(0), R_Y(0)\}$. $R_X = 1$ means $X$ is missing, $R_X = 0$ means $X$ is observed. $R_Y = 1$ means $Y$ is missing,

$R_Y = 0$ means $Y$ is observed. $R_Z = 1$ means $Z$ is missing. $R_Z = 0$ means $Z$ is observed. For $X, Y, Z$, we assume that only one of them is missing at the same time, which means that one of $R_X, R_Y$, and $R_Z$ may be 1, and the other two variables are constant 0.

Ref. [4] defines TBR (treatment benefit rate) as the proportion of the relevant population that benefits from the intervention as compared with the control for a given endpoint. THR (treatment harm rate) is defined as the proportion that is harmed by the intervention as compared with the control based on the same endpoint. Thus, we can use the following equations to describe TBR and THR:

$$TBR = P(Y(0) = 0, Y(1) = 1),$$
$$THR = P(Y(0) = 1, Y(1) = 0).$$

When TBR is much larger than THR, we can say that this treatment is beneficial. On the contrary, when THR is much larger than TBR, we can say that this treatment is harmful.

Let $A \perp\!\!\!\perp B | C$ denote that variables $A$ and $B$ are conditionally independent, given variable $C$. To identify TBR and THR, we need the following assumptions.

**Assumption 1** (Complete randomization). $Z \perp\!\!\!\perp \{Y(0), Y(1), R_X(1), R_X(0), R_Y(1), R_Y(0), R_Z, X\}$.

When the experiment is a completely randomized experiment, its data set is subject to this assumption. It means in a completely randomized experiment, the treatment assignment $Z$ is independent of $\{Y(0), Y(1), R_X(1), R_X(0), R_Y(1), R_Y(0), R_Z, X\}$. This assumption is very strong, and all the theory and methods discussed in this paper are subject to this assumption. Under this assumption, we can get the following equation: $P(Y(z) = 1) = P(Y = 1|Z = z)$.

**Assumption 2.**
$$Y(0) \perp\!\!\!\perp Y(1) | X$$

This assumption means that when the pretreatment covariate $X$ is given, the potential endpoints are independent of each other, that is, given the covariate $X$, $Y(1)$ cannot predict $Y(0)$ and $Y(0)$ cannot predict $Y(1)$.

We aim to identify the TBR and THR through the observed data. If there are no missing data, under the Assumptions 1 and 2, the TBR and THR can be split into the product of two conditional probabilities based on the observed data. Let $TBR_x$ and $THR_x$ denote the treatment benefit rate and treatment harm rate, given $X$. Then, we have:

$$
\begin{aligned}
TBR &= E(TBR_x) = E[P(Y(0) = 0, Y(1) = 1|X)] = E[P(Y(0) = 0|X)P(Y(1) = 1|X)] \\
&= E[P(Y = 0|X, Z = 0)P(Y = 1|X, Z = 1)], \\
THR &= E(THR_x) = E[P(Y(0) = 1, Y(1) = 0|X)] = E[P(Y(0) = 1|X)P(Y(1) = 0|X)] \\
&= E[P(Y = 1|X, Z = 0)P(Y = 0|X, Z = 1)].
\end{aligned}
$$

The above equations illustrate that the TBR and THR can be identified under Assumptions 1 and 2 without missing data. However, when there are missing data in one of the covariate, endpoint, or treatment variables, Assumptions 1 and 2 are not enough to ensure the identification of the TBR and THR, and the above formula no longer works. In this paper, we give sufficient conditions to identify the TBR and THR when one of $X, Y, Z$ have missing data.

Lastly, we introduce the following assumption.

**Assumption 3.** *When $X$ has missing data, $P(R_X = 1|X, Y, Z) < 1$. When $Y$ has missing data, $P(R_Y = 1|X, Y, Z) < 1$. When $Z$ has missing data, $P(R_Z = 1|X, Y, Z) < 1$.*

We need this assumption to ensure that the missing variable is only partially missing.

## 3. Missing Data Mechanisms

In our article, we study the TBR and THR when one of $X$, $Y$, $Z$ have missing data. Before introducing the specific missing mechanisms, we will review the definition of missing at random (MAR) and missing not at random (MNAR) first (Little and Rubin, 2002 [19]). $D_0$ denotes the complete data, $D_1$ denotes the observed data, and $D_2$ denotes the missing data; therefore, we have $D_0 = (D_1, D_2)$. Next, we introduce the two missing mechanisms mentioned above.

**Definition 1.** *The missing data mechanism is called missing at random (MAR), if $R_X$, $R_Y$, or $R_Z$ only depend on the observed data, that is, one of the following three formulas, $R_X \perp\!\!\!\perp D_2|D_1$, $R_Y \perp\!\!\!\perp D_2|D_1$, or $R_Z \perp\!\!\!\perp D_2|D_1$, holds. Otherwise, if $R_X$, $R_Y$, or $R_Z$ depend on $D_2$, the missing data mechanism is called missing not at random (MNAR).*

When the missing indicators only rely on the observed data ($D_1$), because the missing mechanism does not depend on the missing data, the inference for parameters can be based only on the observed data and we call it missing at random (MAR). When the missing is not at random, it is nonignorable. In such a case, we cannot ignore the missing data.

In this article, we study the TBR and THR when one of $X$, $Y$, $Z$ is under the condition of MNAR. The missing mechanisms of $X$, $Y$, or $Z$ are important because they influence the identifiability and estimation of the TBR and THR. For each variable in $X$, $Y$, $Z$, we propose three missing mechanisms.

First, we introduce three missing mechanisms of $X$.

($R_{X1}$) $R_X$ depends on $X$ and $R_X$ is independent of $(Y, Z)$, given $X$, which means:

$$P(R_X = 1|Z = z, X = x, Y = y) = P(R_X = 1|X = x),$$

($R_{X2}$) $R_X$ depends on $(X, Z)$ and $R_X$ is independent of $Y$, given $(X, Z)$, which means:

$$P(R_X = 1|Z = z, X = x, Y = y) = P(R_X = 1|Z = z, X = x), \quad or$$

$$P\{R_X(z) = 1|X = x, Y(z) = y\} = P\{R_X(z) = 1|Z = z, X = x\},$$

($R_{X3}$) $R_X$ depends on $(X, Y)$ and $R_X$ is independent of $Z$, given $(X, Y)$, which means:

$$P(R_X = 1|Z = z, X = x, Y = y) = P(R_X = 1|X = x, Y = y).$$

For the first missing mechanism of $X$, we assume that the missingness of $X$ depends only on $X$. For the second missing mechanism of $X$, we assume that the missingness of $X$ depends on $(X, Z)$. For the third missing mechanism of $X$, we assume that the missingness of $X$ depends on $(X, Y)$. All the missing mechanisms are nonignorable and these missing mechanisms cannot be deduced from each other.

Similarly, we introduce the following several missing mechanisms of $Y$.

($R_{Y1}$) $R_Y$ depends on $Y$ and $R_Y$ is independent of $(X, Z)$, given $Y$, which means:

$$P(R_Y = 1|Z = z, X = x, Y = y) = P(R_Y = 1|Y = y),$$

($R_{Y2}$) $R_Y$ depends on $(Y, Z)$ and $R_Y$ is independent of $X$, given $(Y, Z)$, which means:

$$P(R_Y = 1|Z = z, X = x, Y = y) = P(R_Y = 1|Z = z, Y = y), \quad or$$

$$P\{R_Y(z) = 1|X = x, Y(z) = y\} = P\{R_X(z) = 1|Z = z, Y = y\},$$

($R_{Y3}$) $R_Y$ depends on $(X, Y)$ and $R_Y$ is independent of $Z$, given $(X, Y)$, which means:

$$P(R_Y = 1|Z = z, X = x, Y = y) = P(R_Y = 1|X = x, Y = y).$$

Analogously, for the first missing mechanism of $Y$, we assume that the missingness of $Y$ depends only on $Y$. For the second missing mechanism of $Y$, we assume that the missingness of $Y$ depends on $(Y, Z)$. For the third missing mechanism of $Y$, we assume that the missingness of $Y$ depends on $(X, Y)$. All the missing mechanisms are also nonignorable.

Lastly, we are going to introduce the missing mechanism of $Z$.

$(R_{Z1})$ $R_Z$ depends on $Z$ and $R_Z$ is independent of $(X, Y)$, given $Z$, which means:

$$P(R_Z = 1 | Z = z, X = x, Y = y) = P(R_Z = 1 | Z = z),$$

$(R_{Z2})$ $R_z$ depends on $(Y, Z)$ and $R_Z$ is independent of $X$, which means:

$$P(R_Z = 1 | Z = z, X = x, Y = y) = P(R_Z = 1 | Z = z, Y = y),$$

$(R_{Z3})$ $R_Z$ depends on $(X, Z)$ and $R_Z$ is independent of $Z$, which means:

$$P(R_Z = 1 | Z = z, X = x, Y = y) = P(R_Z = 1 | X = x, Z = z).$$

Above, we introduced three missing mechanisms of $Z$. For the first missing mechanism of $Z$, we assume that the missingness of $Z$ depends only on $Z$. For the second missing mechanism of $Z$, we assume that the missingness of $Z$ depends on $(Y, Z)$. For the third missing mechanism of $Z$, we assume that the missingness of $Z$ depends on $(X, Z)$.

The missing mechanisms mentioned above are all MNAR. The above-mentioned missing mechanisms for $X, Y, Z$ all assume that the missing variable satisfies some conditional independent relationship. In the next section, we consider whether TBR and THR can be identified under these missing mechanisms.

## 4. Identifiability of TBR and THR

In this section, we discuss the identifiability of TBR and THR when one of $X$, $Y$, and $Z$ have missing data. In some mechanisms, we have to identify the joint distribution of $P(X, Y, Z, R_X)$, $P(X, Y, Z, R_Y)$, or $P(X, Y, Z, R_Z)$ to ensure the identifiability of the TBR and THR. We assume the following theorems are under the Assumptions 1 and 2. Before introducing the theorems, note that $X$ has $K$ levels, and $Y$ and $Z$ are both binaries.

Firstly, we give sufficient conditions under which we can identify the TBR and THR when covariate $X$ has missing data.

**Theorem 1.** *For the missing of $X$:*

(1) *Under the missing mechanism $R_{X1}$, the TBR and THR are identifiable when $r(A_1) = r(\bar{A}_1) = K$, where $A_1$ and $\bar{A}_1$ are two matrices and the definitions of $A_1$ and $\bar{A}_1$ are mentioned below, and $r(\cdot)$ is the rank function.*
(2) *Under the missing mechanism $R_{X2}$, the TBR and THR are identifiable when $K = 2$ and $X \not\perp\!\!\!\perp Y | (R_X = 0, Z)$.*
(3) *Under the missing mechanism $R_{X3}$, the TBR and THR are identifiable when $K = 2$ and $X \not\perp\!\!\!\perp Z | (R_X = 0, Y)$.*

When $X$ has missing data, under different missing mechanisms, the identification conditions are also different. Under the first missing mechanism, if we want to identify the THR and THR, we need to assume that the rank of matrix $A_1$ and $\bar{A}_1$ is $K$. $A_1$ and $\bar{A}_1$ are defined as follows.

$$A_1 = \begin{bmatrix} p_{x_1 0 | 00} & p_{x_2 0 | 00} & \cdots & p_{x_K 0 | 00} \\ p_{x_1 0 | 10} & p_{x_2 0 | 10} & \cdots & p_{x_K 0 | 10} \\ p_{x_1 0 | 01} & p_{x_2 0 | 01} & \cdots & p_{x_K 0 | 01} \\ p_{x_1 0 | 11} & p_{x_2 0 | 11} & \cdots & p_{x_K 0 | 11} \end{bmatrix} \text{ and } \bar{A}_1 = \begin{bmatrix} p_{x_1 0 | 00} & p_{x_2 0 | 00} & \cdots & p_{x_K 0 | 00} & 1 \\ p_{x_1 0 | 10} & p_{x_2 0 | 10} & \cdots & p_{x_K 0 | 10} & 1 \\ p_{x_1 0 | 01} & p_{x_2 0 | 01} & \cdots & p_{x_K 0 | 01} & 1 \\ p_{x_1 0 | 11} & p_{x_2 0 | 11} & \cdots & p_{x_K 0 | 11} & 1 \end{bmatrix},$$

where $p_{x_i0|yz} = P(X = x_i, R_X = 0|Y = y, Z = z)$. Note that $A_1$ is a matrix with 4 rows and $K$ columns, and $\bar{A}_1$ is a matrix with 4 rows and $K + 1$ columns. If the rank of $A_1$ and $\bar{A}_1$ is required to be equal to $K$, $K$ must be less than or equal to 4. Under the second missing mechanism, if the covariate and endpoint are not conditionally independent, given $R_X = 0$ and $Z$, and the covariate only has two levels, we can identify the TBR and THR. Under the third missing mechanism, if the covariate and treatment are not conditionally independent, given $R_X = 0$ and $Y$, and the covariate only has two levels, we can also identify the TBR and THR.

Next, we give sufficient conditions under which we can identify the TBR and THR when endpoints $Y$ have missing data.

**Theorem 2.** *For the missing of $Y$:*

*(1) Under the missing mechanism $R_{Y1}$, the TBR and THR are identifiable under the condition $r(B_1) = r(\bar{B}_1) = 2$, where $B_1$ and $\bar{B}_1$ are two matrices and the definitions of $B_1$ and $\bar{B}_1$ can be found in the appendix, and $r(\cdot)$ is the rank function.*

*(2) Under the missing mechanism $R_{Y2}$, the TBR and THR are identifiable under the condition $r(B_{21}) = r(\bar{B_{21}}) = r(B_{22}) = r(\bar{B_{22}}) = 2$, where $B_{21}$, $\bar{B_{21}}$, $B_{22}$, and $\bar{B_{22}}$ are matrices, and the definitions of $B_{21}$, $\bar{B_{21}}$, $B_{22}$, and $\bar{B_{22}}$ can be found in the appendix, and $r(\cdot)$ is the rank function.*

*(3) Under missing mechanism $R_{Y3}$, the TBR and THR are identifiable under the condition $Y \not\perp\!\!\!\perp Z|$ $(R_Y = 0, X = x_i)$ $(i = 1, 2, \ldots, K)$.*

When $Y$ has missing data, we cannot get a uniform identifiable condition. Under different missing mechanisms, it requires different conditions to ensure the identification of the TBR and THR. Under the first missing mechanism, if we want to identify the THR and THR, we need to assume that the rank of the matrix $B_1$ and $\bar{B}_1$ is 2. $B_1$ and $\bar{B}_1$ are defined as follows.

$$B_1 = \begin{bmatrix} p_{00|x_10} & p_{10|x_10} \\ p_{00|x_11} & p_{10|x_11} \\ \vdots & \vdots \\ p_{00|x_K0} & p_{10|x_K0} \\ p_{00|x_K1} & p_{10|x_K1} \end{bmatrix} \text{ and } \bar{B}_1 = \begin{bmatrix} p_{00|x_10} & p_{10|x_10} & 1 \\ p_{00|x_11} & p_{10|x_11} & 1 \\ \vdots & \vdots & \vdots \\ p_{00|x_K0} & p_{10|x_K0} & 1 \\ p_{00|x_K1} & p_{10|x_K1} & 1 \end{bmatrix},$$

where $p_{y0|x_iz} = P(Y = y, R_y = 0|X = x_i, Z = z)$. Under the second missing mechanism, if we want to identify the THR and THR, we need to assume that the rank of matrix $B_{21}$, $\bar{B}_{21}$, $B_{22}$, and $\bar{B}_{22}$ is 2. $B_{21}$, $\bar{B}_{21}$, $B_{22}$, and $\bar{B}_{22}$ are defined as follows.

$$B_{21} = \begin{bmatrix} p_{00|x_10} & p_{10|x_10} \\ \ldots & \ldots \\ p_{00|x_K0} & p_{10|x_K0} \end{bmatrix}, \quad \bar{B}_{21} = \begin{bmatrix} p_{00|x_10} & p_{10|x_10} & 1 \\ \ldots & \ldots \\ p_{00|x_K0} & p_{10|x_K0} & 1 \end{bmatrix},$$

$$B_{22} = \begin{bmatrix} p_{00|x_11} & p_{10|x_11} \\ \ldots & \ldots \\ p_{00|x_K1} & p_{10|x_K1} \end{bmatrix}, \quad \bar{B}_{22} = \begin{bmatrix} p_{00|x_11} & p_{10|x_11} & 1 \\ \ldots & \ldots \\ p_{00|x_K1} & p_{10|x_K1} & 1 \end{bmatrix},$$

where $p_{y0|x_iz} = P(Y = y, R_Y = 0|X = x_i, Z = z)$. Under the last missing mechanism, we can identify the TBR and THR if $Y$ and $Z$ are not conditionally independent, given $R_Y = 0$ and $X = x_i$.

Lastly, we give sufficient conditions under which we can identify the TBR and THR when treatment $Z$ has missing data.

**Theorem 3.** *For the missing of $Z$:*

(1) Under the missing mechanism $R_{Z1}$, the TBR and THR are identifiable under the condition $r(C_1) = r(\bar{C}_1) = 2$, where $C_1$ and $\bar{C}_1$ are two matrices, and the definitions of $C_1$ and $\bar{C}_1$ can be found in the appendix, and $r(\cdot)$ is the rank function.

(2) Under the missing mechanism $R_{Z2}$, the TBR and THR are identifiable under the condition $r(C_{21}) = r(\bar{C}_{21}) = r(C_{22}) = r(\bar{C}_{22}) = 2$, where $C_{21}$, $\bar{C}_{21}$, $C_{22}$, and $\bar{C}_{22}$ are matrices, and the definitions of $C_{21}$, $\bar{C}_{21}$, $C_{22}$, and $\bar{C}_{22}$ can be found in the appendix, and $r(\cdot)$ is the rank function.

(3) Under the missing mechanism $R_{Z3}$, the TBR and THR are identifiable under the condition $Z \not\!\perp\!\!\!\perp Y | X = x_i$ $(i = 1, 2 \ldots K)$.

When $Z$ has missing data, we also cannot get a uniform identifiable condition. Under different missing mechanisms, it requires different conditions to ensure the identification of the TBR and THR. Under the first missing mechanism, if we want to identify the THR and THR, we need to assume that the rank of matrix $C_1$ and $\bar{C}_1$ is 2. $C_1$ and $\bar{C}_1$ are defined as follows.

$$
C_1 = \begin{bmatrix} p_{00|x_10} & p_{10|x_10} \\ p_{00|x_11} & p_{10|x_11} \\ \vdots & \vdots \\ p_{00|x_k0} & p_{10|x_k0} \\ p_{00|x_k1} & p_{10|x_k1} \end{bmatrix} \text{ and } \bar{C}_1 = \begin{bmatrix} p_{00|x_10} & p_{10|x_10} & 1 \\ p_{00|x_11} & p_{10|x_11} & 1 \\ \vdots & \vdots & \vdots \\ p_{00|x_k0} & p_{10|x_k0} & 1 \\ p_{00|x_k1} & p_{10|x_k1} & 1 \end{bmatrix},
$$

where $p_{z0|x_iy} = P(Z = z, R_Z = 0 | X = x_i, Y = y)$. Under the second missing mechanism, if we want to identify the THR and THR, we need to assume that the rank of matrix $C_{21}$, $\bar{C}_{21}$, $C_{22}$, and $\bar{C}_{22}$ is 2. $C_{21}$, $\bar{C}_{21}$, $C_{22}$, and $\bar{C}_{22}$ are defined as follows.

$$
C_{21} = \begin{bmatrix} p_{00|x_10} & p_{10|x_10} \\ \ldots \\ p_{00|x_k0} & p_{10|x_k0} \end{bmatrix}, \quad \bar{C}_{21} = \begin{bmatrix} p_{00|x_10} & p_{10|x_10} & 1 \\ \ldots \\ p_{00|x_k0} & p_{10|x_k0} & 1 \end{bmatrix},
$$

$$
C_{22} = \begin{bmatrix} p_{00|x_11} & p_{10|x_11} \\ \ldots \\ p_{00|x_k1} & p_{10|x_k1} \end{bmatrix}, \quad \bar{C}_{22} = \begin{bmatrix} p_{00|x_11} & p_{10|x_11} & 1 \\ \ldots \\ p_{00|x_k1} & p_{10|x_k1} & 1 \end{bmatrix},
$$

where $p_{z0|x_iy} = P(Z = z, R_Z = 0 | X = x_i, Y = y)$. Under the last missing mechanism, we can identify the TBR and THR if $Y$ and $Z$ are not conditionally independent, given $R_Y = 0$ and $X = x_i$.

The above three theorems give sufficient conditions under which the TBR and THR can be identified. In the next two parts, we illustrate my conclusion through simulation and actual data.

## 5. Computational Details and Simulation Study

In this part, we first introduce how to use the EM algorithm to estimate the TBR and THR when covariate $X$ has missing data and satisfies missing mechanism $R_{X2}$. When $X$ satisfies other missing mechanisms or there are missing data in the other two variables, the estimation is similar. Next, we generate simulation data and then apply our method to the simulation data to illustrate that our estimation works. We use statistical software R to implement our numerical simulation.

### 5.1. Expectation Maximization Algorithms

We define $P_{xyzr} = P(X = x, Y = y, Z = z, R_X = r)$ and $P_{+yzr} = P(Y = y, Z = z, R_X = r)$, where "+" represents the marginal distribution over corresponding variable. Similarly, let $N_{xyzr}$ denote the observed frequency in the cell $(x, y, z, r)$ of the contingency table, and $N_{+yzr}$ denote the marginal frequency of the contingency table over the corresponding variable $X$. When "+" is at another position, its meaning is the same.

In practice, we can use the expectation maximization (EM) algorithm to find the MLEs. In this subsection, we only describe the computational details for missing mechanism $R_{X2}$. For simplicity, we only describe the algorithms for binary $X$. The algorithms for multi-categorical $X$ can be written similarly. Under the missing mechanism $R_{X2}$, we have $R_X \perp\!\!\!\perp Y | (X, Z)$. Thus, the joint distribution of $(X, Y, Z, R_X)$ can be written as $P^{(j)}(X = x, Y = y, Z = z, R_X = r) = P^{(j)}(X = x)P^{(j)}(Z = z | X = x)P^{(j)}(Y = y | X = x, Z = z)P^{(j)}(R_X = r | X = x, Z = z)$. Superscript $j$ indicates the $j$-th iteration. Define:

$$
\begin{aligned}
p^{(j)}_{x|yz1} &= p^{(j)}(X = x | Y = y, Z = z, R_X = 1) \\
&= \frac{P^{(j)}(X = x, Y = y, Z = z, R_X = 1)}{\sum_{x'=0,1} P^{(j)}(X = x', Y = y, Z = z, R_X = 1)}
\end{aligned}
$$

The EM algorithm iterates between the following E-step and M-step:

(a)   E-step: The sufficient statistics are imputed as $N^{(j)}_{xyz0} = N_{xyz0}$ and $N^{(j)}_{xyz1} = N_{+yz1} p^{(j)}_{x|yz1}$;

(b)   M-step: The joint distribution is updated by $P^{(j+1)}(X = x, Y = y, Z = z, R_X = 1) = \frac{N^{(j)}_{x+++}}{N^{(j)}_{++++}} \frac{N^{(j)}_{x+z+}}{N^{(j)}_{x+++}} \frac{N^{(j)}_{xyz+}}{N^{(j)}_{x+z+}} \frac{N^{(j)}_{x+zr}}{N^{(j)}_{x+z+}}.$

After the algorithm converges, we assume that the convergent probability is $\hat{P}$. According to the formula in the second section, we can estimate the TBR and THR as follows.

$$
\begin{aligned}
\text{TBR} &= \hat{P}(Y = 0 | X = 1, Z = 0)\hat{P}(Y = 1 | X = 1, Z = 1)\hat{P}(X = 1) \\
&+ \hat{P}(Y = 0 | X = 0, Z = 0)\hat{P}(Y = 1 | X = 0, Z = 1)\hat{P}(X = 0), \\
\text{THR} &= \hat{P}(Y = 1 | X = 1, Z = 0)\hat{P}(Y = 0 | X = 1, Z = 1)\hat{P}(X = 1) \\
&+ \hat{P}(Y = 1 | X = 0, Z = 0)\hat{P}(Y = 0 | X = 0, Z = 1)\hat{P}(X = 0).
\end{aligned}
$$

Lastly, we calculate the standard errors of the above estimator by repeating the processes 1000 times.

### 5.2. Simulation Study

In this section, we evaluate the finite sample performances of the likelihood-based estimator for the missing mechanisms $R_{X1}$ and $R_{Y2}$ via simulation studies. In order to mimic the real data analyzed in the next section, we assume that $Z$ is completely randomized and $Z \perp\!\!\!\perp X$. We generated $Z \sim Bernoulli(0.5)$ and $X \sim Bernoulli(0.5)$. $P_{y|xz} = P(Y = y | Z = z, X = x)$ is defined, and $Y$ is generated according to the conditional distribution $(P_{1|00}, P_{1|10}, P_{1|01}, P_{1|11})$. We set the parameters of the two missing mechanisms as follows.

($R_{X2}$):
$$(P_{1|00}, P_{1|01}, P_{1|10}, P_{1|11}) = (0.4, 0.5, 0.8, 0.4),$$
$$P(R_X = 0 | X = 1, Z = 1) = 0.7, P(R_X = 0 | X = 0, Z = 1) = 0.4,$$
$$P(R_X = 0 | X = 1, Z = 0) = 0.5, P(R_X = 0 | X = 0, Z = 0) = 0.6.$$

($R_{Y3}$):
$$(P_{1|00}, P_{1|01}, P_{1|10}, p_{1|11}) = (0.4, 0.5, 0.5, 0.6),$$
$$P(R_Y = 0 | X = 1, Y = 1) = 0.4, P(R_Y = 0 | X = 0, Y = 1) = 0.3,$$
$$P(R_Y = 0 | X = 1, Y = 1) = 0.4, P(R_Y = 0 | X = 0, Y = 1) = 0.3.$$

We use the EM algorithm to find the MLEs of the parameters and calculate the corresponding THR and TBR. The sample sizes of the simulation study are 500, 1000, and 1500, respectively, and we repeat the simulation 1000 times. The means and the standard errors of the estimates of the TBR and THR are given in Tables 1 and 2.

**Table 1.** TBR and THR mean(sd) by 10,000 repetitions (under the missing mechanism $R_{X2}$).

| True Value | Sample Size n | | |
|---|---|---|---|
| | 500 | 1000 | 1500 |
| TBR = 0.41 | 0.4071 (0.0282) | 0.4082 (0.0204) | 0.4086 (0.0156) |
| THR = 0.11 | 0.1138 (0.0146) | 0.1135 (0.0104) | 0.1134 (0.0081) |

**Table 2.** TBR and THR mean(sd) by 10,000 repetitions (under the missing mechanism $R_{Y3}$).

| True Value | Sample Size n | | |
|---|---|---|---|
| | 500 | 1000 | 1500 |
| TBR = 0.30 | 0.2900 (0.0228) | 0.2900 (0.0155) | 0.2905 (0.0135) |
| THR = 0.21 | 0.2132 (0.1976) | 0.2130 (0.01333) | 0.2125 (0.0115) |

We can see from the simulation results that the values of TBR and THR can be estimated consistently, which means that the TBR and THR are identifiable. With the increase of sample size, the standard deviation decreases gradually.

## 6. Application

In this part, we illustrate the correctness of our method with three real data examples.

### 6.1. Application to the Second Multicenter Automatic Defibrillator Intervention Trial

In this section, we re-analyzed a randomized clinical trial using the newly proposed methods under the missing mechanism $R_{X2}$. We first briefly review the background of the illustrative clinical trial, and more details of the data can be found in the previous paper ([8]). In this example, $Z$ is the treatment assignment variable, with $Z = 1$ denoting the treatment (implantable cardiac defibrillator) and $Z = 0$ denoting the control. The endpoint $Y$ is the death indicator, with $Y = 1$ denoting dead and $Y = 0$ denoting alive. Let $X$ denote the inducible indicator, with $X = 1$ denoting inducible and $X = 0$ denoting noninducible. The covariate $X$ is obtained from the electro-physiological stimulation (EPS) testing. Because the EPS testing is invasive and not a pre-requisite for enrollment in the study, 79.3% of patients in the implantable cardiac defibrillator arm have EPS records, whereas only 24% of patients in the control arm have EPS records. Therefore, the problem of missing data for the covariate $X$ is very severe. The observed data can be summarized as the following counts ($N_{xyzr_x}$): $N_{0000} = 4$, $N_{0010} = 311$, $N_{1000} = 6$, $N_{1010} = 190$, $N_{0100} = 0$, $N_{0110} = 62$, $N_{1100} = 2$, $N_{1110} = 20$, $N_{+001} = 382$, $N_{+101} = 95$, $N_{+011} = 136$, and $N_{+111} = 23$. We assume that the missing mechanism of $X$ is $R_{X2}$. Firstly, we use the EM algorithm to calculate the maximum likelihood estimation of the parameters and then calculate the *TBR* and *THR*. Then, the sampling is repeated 1500 times to calculate the standard deviation of TBR and THR. The estimated TBR and THR are 0.1121 (0.0121) and 0.1697 (0.0175). The numbers in brackets indicate the standard deviation.

### 6.2. Application to the Mechanical Treatment Trial for Crisis Patients

In this section, we will re-analyze a randomized clinical trial using the newly proposed methods under missing mechanism $R_{Y3}$. We first briefly review the background of the trial ([9]). In this example, critically ill patients randomly received mechanical ventilation 1:1 within each study site to manage with a paired sedation plus ventilator weaning protocol involving the daily interruption of sedative through spontaneous awakening trials (SATs) and spontaneous breathing trials (SBTs) or sedation per usual care (UC) and SBTs. $Z$ is the treatment assignment variable, with $Z = 1$ denoting the treatment (SAT and SBT) and $Z = 0$ denoting the control (UC and SBT). The endpoint $Y$ is the cognitive score, with $Y = 1$ denoting "higher cognitive ability" and $Y = 0$ denoting "lower cognitive ability". Let $X$ denote age, with $X = 1$ denoting "the people older than 33 years old" and

$X = 0$ denoting "the people younger than 33 years old". In randomized studies involving severely ill patients, functional endpoints are often unobserved due to missed clinic visits, premature withdrawal, or death. The observed data can be summarized as the following counts ($N_{xyzr_y}$): $N_{0000} = 9$, $N_{0100} = 9$, $N_{1000} = 7$, $N_{1100} = 0$, $N_{0010} = 12$, $N_{0110} = 16$, $N_{1010} = 10$, $N_{1110} = 6$, $N_{0+01} = 24$, $N_{1+01} = 45$, $N_{0+11} = 23$, and $N_{1+11} = 26$. We assume that the missing mechanism of $Y$ is $R_{Y3}$. Similarly, we use the EM algorithm to calculate the maximum likelihood estimation of the parameters and then calculate the $TBR$ and $THR$. Then, we use the bootstrap method to repeat sampling 1500 times to calculate the standard deviation of the TBR and THR. The estimated TBR and THR are 0.270 (0.043) and 0.230 (0.039). The numbers in brackets indicate the standard deviation.

*6.3. Application to the Job Search Intervention Study*

In this section, we will analyze a randomized trial using the proposed methods under missing mechanism $R_{Z1}$. Firstly, we will introduce the background of the data. The Job Search Intervention Study (JOBS II) was a randomized field experiment that investigated the efficacy of a job training intervention on unemployed workers ([20]). There are 899 unemployed workers in the "jobs" dataset. All the workers were randomly assigned to two groups, the control group (people received a booklet describing job-search process) and the treatment group (people participated in job skills workshops); the binary endpoint represents whether the respondents had become employed. $Z$ is the treatment assignment variable, with $Z = 1$ denoting the treatment (people participated in job skills workshops) and $Z = 0$ denoting the control (people received a booklet describing job search process). $Y$ denotes the endpoint; $Y = 1$ denotes that the worker became employed finally, that is, the treatment worked; and $Y = 0$ denotes that the worker was still unemployed. Additionally, $X$ denotes sex, with $X = 0$ for female and $X = 1$ for male. The observed data can be summarized as the following counts ($N_{xyz}$): $N_{111} = 211$, $N_{011} = 182$, $N_{101} = 99$, $N_{100} = 38$, $N_{110} = 134$, $N_{010} = 79$, $N_{000} = 48$, $N_{001} = 108$. Based on this data, we assume $P(R_Z = 1|Z = 1) = 0.3$, $P(R_Z = 1|Z = 0) = 0.2$ and manually generate missing data. The generated data can be summarized as the following counts ($N_{xyzr_z}$): $N_{00+1} = 47$, $N_{01+1} = 63$, $N_{10+1} = 41$, $N_{11+1} = 95$, $N_{0000} = 35$, $N_{0100} = 68$, $N_{1000} = 24$, $N_{1100} = 102$, $N_{0010} = 74$, $N_{0110} = 130$, $N_{1010} = 72$, $N_{1110} = 148$. Similarly, we use the EM algorithm to calculate the maximum likelihood estimation of the parameters and then calculate the $TBR$ and $THR$. We use the bootstrap method to repeat the sampling 1500 times to calculate the standard deviation of the TBR and THR. The estimated TBR and THR are 0.2090 (0.0160) and 0.1870 (0.0163). The number in brackets indicates the standard deviation.

## 7. Discussion

In the field of causal inference, the average causal effect is an important measure, but this measure is also flawed. Its flaw is that it ignores the heterogeneous responses to the treatment in the target population. Therefore, in this article, we study the TBR and THR proposed by [4]. In addition, in randomized experiments, the existence of missing data is a common phenomenon [21], so we assume that there are missing data in one of the covariate, endpoint, or treatment. We give sufficient conditions to make the TBR and THR identifiable in the presence of missing data. We illustrate our method through simulated data, and then apply our method to several actual data.

There are several issues beyond the scope of this paper. First, in Assumption 2, we assume that given a covariate $X$, the two potential variables $Y_0$ and $Y_1$ are conditionally independent. This assumption also appeared in [4]. However, we can only observe one of the two potential variables, and the other one cannot be observed, which means that Assumption 2 cannot be verified by the data. Thus, it is better to propose a more appropriate assumption to ensure that TBR and THR can be identified.

Second, in our article, we assume that the covariate $X$ in Assumption 1 is a binary one-dimensional variable. However, in practice, $X$ may be a continuous variable or high-

dimensional variable, and there may also be unobservable variables in $X$. In this case, even if there are no missing data, it is very difficult to identify the TBR and THR because the observations in each subgroup may be very sparse in a limited sample. If there are still missing data, we need to propose new conditions so that the TBR and THR can be identified.

Third, we discussed the situation where only one of the covariate, endpoint, and treatment variables may be MNAR. In many applications, both the covariate and the endpoint may be MNAR at the same time. In this case, the identification and estimation of the TBR and THR will be more complicated.

Although the problems mentioned above are beyond the scope of this article, we will continue our research in this area.

**Author Contributions:** Conceptualization, P.L.; methodology, Y.H.; software, L.Z.; formal analysis, L.Z.; investigation, Y.H.; writing—original draft preparation, Y.H.; writing—review and editing, P.L.; supervision, P.L.; project administration, P.L. All authors have read and agreed to the published version of the manuscript.

**Data Availability Statement:** The data was found in the R software package.

**Conflicts of Interest:** The authors declare no conflict of interest.

## Appendix A. Proofs of Theorems

Note that:

$$TBR = E_X[Pr(Y(0) = 0, Y(1) = 1|X)] = E_x[Pr(Y = 0|X, Z = 0)Pr(Y = 1|X, Z = 1)]$$

$$THR = E_X[Pr(Y(0) = 1, Y(1) = 0|X)] = E_x[Pr(Y = 1|X, Z = 0)Pr(Y = 0|X, Z = 1)]$$

From the above formula, we can see that as long as we identify $P(Y|X, Z)$ and $P(X)$, we can identify the TBR and THR. As such, in the following proof, the primary task is to prove that $P(Y|X, Z)$ and $P(X)$ are identifiable. In our proof, we regard uppercase letters as random variables and lowercase letters as the values of random variables.

**Proof of Theorem 1.** We will use the following notation in proof 1:
$p_{xr_x|yz} = P(X = x, R_X = r_x|Y = y, Z = z)$, $p_{r_x|x} = P(R_X = r_x|X = x)$, $p_{r_x|xz} = P(R_X = r_x|X = x, Z = z)$, $p_{r_x|xy} = P(R_X = r_x|X = x, Y = y)$

(1): Because $(Y, Z) \perp\!\!\!\perp R_X|X$, we have $P(X, Y, Z, R_X) = P(Y, Z)P(X|Y, Z)P(R_X|X, Y, Z)$ $= P(Y, Z)P(X|Y, Z)P(R_X|X)$. Divide both sides by $P(Y, Z)P(R_X|X)$ and consider the missing mechanism, and we can get:

$$\frac{P(X, R_X = 0|Y, Z)}{P(R_X = 0|X)} = P(X|Y, Z),$$

and

$$\sum_{i=1}^{K} \frac{P(X = x_i, R_X = 0|Y, Z)}{P(R_X = 0|X = x_i)} = \sum_{i=1}^{K} P(X = x_i|Y, Z) = 1.$$

To identify $P(X)$ and $P(Y|X, Z)$, we have to identify $P(X, Y, Z, R_X)$ first, and calculate the marginal distribution of $X$ and the conditional distribution of $Y$ after, given $X, Z$. In the equation mentioned above, $P(Y, Z)$ is identifiable, so we have to prove that $P(X|Y, Z)$ and

$P(R_X = 1|X, Z)$ are also identifiable. Write the above formula in the form of matrix, and then we can get the following equations:

$$
\begin{bmatrix}
p_{x_1 0|00} & p_{x_2 0|00} & \cdots & p_{x_K 0|00} \\
p_{x_1 0|10} & p_{x_2 0|10} & \cdots & p_{x_K 0|10} \\
p_{x_1 0|01} & p_{x_2 0|01} & \cdots & p_{x_K 0|01} \\
p_{x_1 0|11} & p_{x_2 0|11} & \cdots & p_{x_K 0|11}
\end{bmatrix}
\begin{bmatrix}
\frac{1}{p_{0|x_1}} \\
\frac{1}{p_{0|x_2}} \\
\cdots \\
\frac{1}{p_{0|x_K}}
\end{bmatrix}
=
\begin{bmatrix}
1 \\ 1 \\ 1 \\ 1
\end{bmatrix}.
$$

To identify $P(X|Y, Z)$, the above equation should have only one solution. Notice that the equation is a linear equation and has $K$ unknowns; they are $\frac{1}{p_{0|x_1}}$, $\frac{1}{p_{0|x_2}}$, $\ldots$ and $\frac{1}{p_{0|x_2}}$. Suppose the coefficient matrix and augmented matrix of this equation are $A_1$ and $\bar{A}_1$, respectively, which means:

$$
A_1 =
\begin{bmatrix}
p_{x_1 0|00} & p_{x_2 0|00} & \cdots & p_{x_K 0|00} \\
p_{x_1 0|10} & p_{x_2 0|10} & \cdots & p_{x_K 0|10} \\
p_{x_1 0|01} & p_{x_2 0|01} & \cdots & p_{x_K 0|01} \\
p_{x_1 0|11} & p_{x_2 0|11} & \cdots & p_{x_K 0|11}
\end{bmatrix}
\text{ and } \quad
\bar{A}_1 =
\begin{bmatrix}
p_{x_1 0|00} & p_{x_2 0|00} & \cdots & p_{x_K 0|00} & 1 \\
p_{x_1 0|10} & p_{x_2 0|10} & \cdots & p_{x_K 0|10} & 1 \\
p_{x_1 0|01} & p_{x_2 0|01} & \cdots & p_{x_K 0|01} & 1 \\
p_{x_1 0|11} & p_{x_2 0|11} & \cdots & p_{x_K 0|11} & 1
\end{bmatrix}
$$

According to the theory of linear equations, when the rank of the coefficient matrix, the rank of the augmented matrix, and the number of unknowns are the same($r(A) = r(\bar{A}) = K$), the equation has a unique solution. After identifying $P(R_X|X)$, according to the above formula, $P(X|Y, Z)$ can also be identified, and then the entire joint distribution can be identified. As such, the TBR and THR can also be identified.

(2): Because $Y \perp\!\!\!\perp R_x|(X, Z)$, we have $P(Y|X, Z) = P(Y|X, Z, R_X = 0)$. Notice that $P(Y|X, Z, R_X = 0)$ is identifiable; thus, $P(Y|X, Z)$ is identifiable. In addition, for P(X), notice that $P(X, Y, Z, R_X) = P(Y, Z)P(X|Y, Z)P(R_X|X, Y, Z) = P(Y, Z)P(X|Y, Z)P(R_X|X, Z)$. Divide both sides by $P(Y, Z)P(R_X|X, Y, Z)$ and consider the missing mechanism, and we can get:

$$
\frac{P(X, R_X = 0|Y, Z)}{P(R_X = 0|X, Z)} = P(X|Y, Z),
$$

and

$$
\sum_{i=1}^{K} \frac{P(X = x_i, R_X = 0|Y, Z)}{P(R_X = 0|X = x_i, Z)} = \sum_{i=1}^{K} P(X = x_i|Y, Z) = 1.
$$

To identify $P(X)$, we have to identify $P(X, Y, Z, R_X)$ and calculate its marginal distribution with respect to $X$. In the equation mentioned above, $P(Y, Z)$ is identifiable, and so we have to prove that $P(X|Y, Z)$ and $P(R_X = 0|X, Z)$ are identifiable. According to the above equations, we can get:

$$
\begin{bmatrix}
p_{x_1 0|00} & p_{x_2 0|00} & \cdots & p_{x_K 0|00} \\
p_{x_1 0|10} & p_{x_2 0|10} & \cdots & p_{x_K 0|10}
\end{bmatrix}
\begin{bmatrix}
\frac{1}{p_{0|x_1 0}} \\
\frac{1}{p_{0|x_2 0}} \\
\cdots \\
\frac{1}{p_{0|x_K 0}}
\end{bmatrix}
=
\begin{bmatrix}
1 \\ 1
\end{bmatrix}
$$

$$
\begin{bmatrix}
p_{x_1 0|01} & p_{x_2 0|01} & \cdots & p_{x_K 0|01} \\
p_{x_1 0|11} & p_{x_2 0|11} & \cdots & p_{x_K 0|11}
\end{bmatrix}
\begin{bmatrix}
\frac{1}{p_{0|x_1 1}} \\
\frac{1}{p_{0|x_2 1}} \\
\cdots \\
\frac{1}{p_{0|x_K 1}}
\end{bmatrix}
=
\begin{bmatrix}
1 \\ 1
\end{bmatrix}.
$$

To identify $P(X|Y,Z)$, the above equations should have only one solution. Notice that both of these equations have $K$ unknowns, and the rank of coefficient matrix is less than or equal to 2, so to have a unique solution we need $K \leq 2$. In addition, we have mentioned before that $K \geq 2$, so there is a unique solution if $K = 2$.

When $K = 2$, if the solution to the equation is unique, the coefficient matrix must be invertible. $\Leftrightarrow p_{x_10|00}p_{x_20|10} \neq p_{x_20|00}p_{x_10|10}$ and $p_{x_10|01}p_{x_20|11} \neq p_{x_10|11}p_{x_20|01}$. $\Leftrightarrow X \perp\!\!\!\perp Y|(R_X = 0, Z)$.

(3): Because of the missing mechanism $R_X \perp\!\!\!\perp Z|(X,Y)$, we can spilt the joint distribution: $P(X,Y,Z,R_X) = P(Y,Z)P(X|Y,Z)P(R_X|X,Y,Z) = P(Y,Z)P(X|Y,Z)P(R_X|X,Y)$. Divide both sides by $P(Y,Z)P(R_X|X,Y)$, and we can get:

$$\frac{P(X, R_X = 0|Y,Z)}{P(R_X = 0|X,Y)} = P(X|Y,Z)$$

and

$$\sum_{i=1}^{K} \frac{P(X = x_i, R_X = 0|Y,Z)}{P(R_X = 0|X,Y)} = \sum_{i=1}^{K} P(X = x_i|Y,Z) = 1.$$

To identify $P(Y|X,Z)$ and $P(X)$, we should identify $P(X,Y,Z,R_X)$ first. Similarly, write the above equation in the form of matrix, and we can get the following equations:

$$\begin{bmatrix} p_{x_10|00} & p_{x_20|00} & \cdots & p_{x_K0|00} \\ p_{x_10|01} & p_{x_20|01} & \cdots & p_{x_K0|01} \end{bmatrix} \begin{bmatrix} \frac{1}{p_{0|x_10}} \\ \frac{1}{p_{0|x_20}} \\ \cdots \\ \frac{1}{p_{0|x_K0}} \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

$$\begin{bmatrix} p_{x_10|10} & p_{x_20|10} & \cdots & p_{x_K0|10} \\ p_{x_10|11} & p_{x_20|11} & \cdots & p_{x_K0|11} \end{bmatrix} \begin{bmatrix} \frac{1}{p_{0|x_11}} \\ \frac{1}{p_{0|x_21}} \\ \cdots \\ \frac{1}{p_{0|x_K1}} \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

To identify $P(X,Y,Z,R_X)$, the above equations should have only one solution. Notice that both of these equations have $K$ unknowns, and the rank of the equation coefficient matrix is less than or equal to 2. In order to have a unique solution, we need the condition that $K$ is less than or equal to 2. We have mentioned that $K \geq 2$, so there is a unique solution if, and only if, $K$ is equal to 2.

When $K = 2$, the solution to the equation is unique. $\Leftrightarrow$ The coefficient matrix is invertible. $\Leftrightarrow P_{x_10|00}P_{x_20|01} \neq P_{x_20|00}P_{x_10|01}$ and $P_{x_10|10}P_{x_20|11} \neq P_{x_10|11}P_{x_20|10}$ are true. $\Leftrightarrow X \perp\!\!\!\perp Z|(R_X = 0, Y)$. $\square$

**Proof of Theorem 2.** We will use the following notation in proofs 2:
$p_{yr_y|xz} = P(Y = y, R_Y = r_y|X = x, Z = z)$, $p_{r_y|y} = P(R_Y = r_y|Y = y)$, $p_{r_y|yz} = P(R_Y = r_y|Y = y, Z = z)$, $p_{r_y|xy} = P(R_Y = r_y|X = x, Y = y)$

(1) We can spilt the joint distribution: $P(X,Y,Z,R_Y) = P(X,Z)P(Y|X,Z)P(R_Y|X,Y,Z)$. Divide both sides by $P(X,Z)P(R_Y|X,Y,Z)$, because of the missing mechanism $R_Y \perp\!\!\!\perp (X,Z)|Y$, and we can get:

$$\frac{P(Y, R_Y = 0|X,Z)}{P(R_Y = 0|Y)} = P(Y|X,Z),$$

and

$$\sum_{y=0,1} \frac{P(Y = y, R_Y = 0|X,Z)}{P(R_Y = 0|Y)} = \sum_{y=0,1} P(Y = y|X,Z) = 1.$$

$P(X, Z)$ is identifiable, and to identify $P(X, Y, Z, R_Y)$, we need to prove that $P(Y|X, Z)$ and $P(R_Y = 0|Y)$ are identifiable. Write the above formula in the form of matrix, and we can get:

$$
\begin{bmatrix}
p_{00|x_10} & p_{10|x_10} \\
p_{00|x_11} & p_{10|x_11} \\
\cdots & \cdots \\
p_{00|x_K0} & p_{10|x_K0} \\
p_{00|x_K1} & p_{10|x_K1}
\end{bmatrix}
\begin{bmatrix}
\frac{1}{p_{0|0}} \\
\frac{1}{p_{0|1}}
\end{bmatrix}
=
\begin{bmatrix}
1 \\
1 \\
\cdots \\
1 \\
1
\end{bmatrix},
$$

To identify the unknown parameters, the equations should only have one solution. Thus, the rank of coefficient matrix and augmented matrix is 2. Suppose the coefficient matrix and augmented matrix of this equation are $B_1$ and $\bar{B}_1$, respectively, which means:

$$
B_1 =
\begin{bmatrix}
p_{00|x_10} & p_{10|x_10} \\
p_{00|x_11} & p_{10|x_11} \\
. & \\
. & \\
. & \\
p_{00|x_K0} & p_{10|x_K0} \\
p_{00|x_K1} & p_{10|x_K1}
\end{bmatrix}
\text{ and } \bar{B}_1 =
\begin{bmatrix}
p_{00|x_10} & p_{10|x_10} & 1 \\
p_{00|x_11} & p_{10|x_11} & 1 \\
. & & \\
. & & \\
. & & \\
p_{00|x_K0} & p_{10|x_K0} & 1 \\
p_{00|x_K1} & p_{10|x_K1} & 1
\end{bmatrix}
$$

According to the theory of linear equations, when the rank of the coefficient matrix, the rank of the augmented matrix, and the number of unknowns are the same ($r(B_1) = r(\bar{B}_1) = 2$), the equation has a unique solution.

(2) We can spilt the joint distribution: $P(X, Y, Z, R_Y) = P(X, Z)P(Y|X, Z)P(R_Y|X, Y, Z)$. Divide both sides by $P(X, Z)P(R_Y|X, Y, Z)$ and consider the missing mechanism $R_Y \perp\!\!\!\perp X|(Y, Z)$, and we can get:

$$
\frac{P(Y, R_Y = 0|X, Z)}{P(R_Y = 0|Y, Z)} = P(Y|X, Z),
$$

and

$$
\sum_{y=0,1} \frac{P(Y = y, R_Y = 0|X, Z)}{P(R_Y = 0|Y, Z)} = \sum_{y=0,1} P(Y = y|X, Z) = 1.
$$

Because $P(X, Z)$ is identifiable, to identify $P(X, Y, Z, R_Y)$, we need to prove that $P(Y|X, Z)$ and $P(R_Y = 1|Y, Z)$ are identifiable. Write the above equations in the form of a matrix, and we can get:

$$
\begin{bmatrix}
p_{00|x_10} & p_{10|x_10} \\
\cdots & \cdots \\
p_{00|x_K0} & p_{10|x_K0}
\end{bmatrix}
\begin{bmatrix}
\frac{1}{p_{0|00}} \\
\frac{1}{p_{0|10}}
\end{bmatrix}
=
\begin{bmatrix}
1 \\
\cdots \\
1
\end{bmatrix},
\tag{A1}
$$

$$
\begin{bmatrix}
p_{00|x_11} & p_{10|x_11} \\
\cdots & \cdots \\
p_{00|x_K1} & p_{10|x_K1}
\end{bmatrix}
\begin{bmatrix}
\frac{1}{p_{0|01}} \\
\frac{1}{p_{0|11}}
\end{bmatrix}
=
\begin{bmatrix}
1 \\
\cdots \\
1
\end{bmatrix}.
\tag{A2}
$$

To identify the unknown parameters, the equations should only have one solution. Thus, the rank of coefficient matrices and augmented matrices is 2. Suppose the coefficient matrix and augmented matrix of this equation are $B_{21}$, $B_{22}$, and $\bar{B}_{21}$, $\bar{B}_{22}$, respectively, which means:

$$
B_{21} =
\begin{bmatrix}
p_{00|x_10} & p_{10|x_10} \\
\cdots & \\
p_{00|x_K0} & p_{10|x_K0}
\end{bmatrix},
\ \bar{B}_{21} =
\begin{bmatrix}
p_{00|x_10} & p_{10|x_10} & 1 \\
\cdots & & \\
p_{00|x_K0} & p_{10|x_K0} & 1
\end{bmatrix},
$$

$$B_{22} = \begin{bmatrix} p_{00|x_1 1} & p_{10|x_1 1} \\ \ldots & \\ p_{00|x_K 1} & p_{10|x_K 1} \end{bmatrix}, \quad \bar{B_{22}} = \begin{bmatrix} p_{00|x_1 1} & p_{10|x_1 1} & 1 \\ \ldots & & \\ p_{00|x_K 1} & p_{10|x_K 1} & 1 \end{bmatrix},$$

According to the theory of linear equation, when the rank of the coefficient matrices, the rank of the augmented matrices, and the number of unknowns are the same $(r(B_{21}) = r(\bar{B_{21}}) = r(B_{22}) = r(\bar{B_{22}}) = 2)$, the equations have a unique solution.

(3) We can split the joint distribution: $P(X, Y, Z, R_Y) = P(X, Z)P(Y|X, Z)P(R_Y|X, Y)$. Divide both sides by $P(X, Z)P(R_Y|X, Y)$ and consider the missing mechanism $R_Y \perp\!\!\!\perp Z | (X, Y)$, and we can get:

$$\frac{P(Y, R_Y = 0|X, Z)}{P(R_Y = 0|X, Y)} = P(Y|X, Z),$$

and

$$\sum_{y=0,1} \frac{P(Y = y, R_Y = 0|X, Z)}{P(R_Y = 0|X, Y = y)} = \sum_{y=0,1} P(Y = y|X, Z) = 1.$$

$P(X, Z)$ is identifiable, and to identify $P(X, Y, Z, R_Y)$, we need to prove that $P(R_Y = 0|X, Y)$ and $P(Y|X, Z)$ are identifiable. Similarly, we can get following equations:

$$\begin{bmatrix} p_{00|x_1 0} & p_{10|x_1 0} \\ p_{00|x_1 1} & p_{10|x_1 1} \end{bmatrix} \begin{bmatrix} \frac{1}{p_{0|x_1 0}} \\ \frac{1}{p_{0|x_1 1}} \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \tag{A3}$$

$$\ldots$$

$$\begin{bmatrix} p_{00|x_k 0} & p_{10|x_k 0} \\ p_{00|x_k 1} & p_{10|x_k 1} \end{bmatrix} \begin{bmatrix} \frac{1}{p_{0|x_k 0}} \\ \frac{1}{p_{0|x_k 1}} \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}. \tag{A4}$$

To identify the unknown parameters, the equations should only have one solution; it is equivalent to:

$$\begin{vmatrix} p_{00|x_i 0} & p_{10|x_i 0} \\ p_{00|x_i 1} & p_{10|x_i 1} \end{vmatrix} \neq 0 \quad i = 1, 2 \ldots k,$$

$$\Leftrightarrow \frac{p_{00|x_i 0}}{p_{00|x_i 0}} \neq \frac{p_{10|x_i 1}}{p_{10|x_i 1}}$$

$$\Leftrightarrow Y \not\perp\!\!\!\perp Z | (R_Y = 0, X = x_i) (i = 1, 2 \ldots k)$$

In sum, if the condition $Y \not\perp\!\!\!\perp Z | (R_Y = 0, X = x_i) (i = 1, 2 \ldots k)$ is satisfied, we can identify joint distribution $P(X, Y, Z, R_Y)$, and then we can identify the TBR and THR naturally. □

**Proof of Theorem 3.** We will use the following notations in the proof:
$p_{zr_z|xy} = P(Z = z, R_Z = r_z|X = x, Y = y)$, $p_{r_z|z} = P(R_Z = r_z|Z = z)$, $p_{r_z|xz} = P(R_Z = r_z|X = x, Z = z)$, $p_{r_z|yz} = P(R_Z = r_z|Y = y, Z = z)$.

(1): the missing mechanism is $(X, Y) \perp\!\!\!\perp R_Z | Z$, thus

$$P(X, Y, Z, R_Z) = P(X, Y)P(Z|X, Y)P(R_Z|X, Y, Z) = P(X, Y)P(Z|X, Y)P(R_Z|Z).$$

Divide both sides by $P(X, Y)P(R_Z|Z)$, and we can get:

$$\frac{P(Z, R_Z = 0|X, Y)}{P(R_Z = 0|Z)} = P(Z|X, Y),$$

and

$$\sum_{z=0,1} \frac{P(Z=z, R_Z=0|X,Y)}{P(R_Z=0|Z=z)} = \sum_{z=0,1} P(Z=z|X,Y) = 1.$$

To identify $P(X)$ and $P(Y|X,Z)$, we have to identify $P(X,Y,Z,R_Z)$ first, and then calculate the marginal distribution with respect to $X$ and calculate the conditional distribution of $Z$ after, given $X,Y$. In the equation mentioned above, $P(X,Y)$ is identifiable, so we have to prove that $P(Z|X,Y)$ and $P(R_Z=0|X,Z)$ are identifiable. Similarly, we can get:

$$\begin{bmatrix} p_{00|x_10} & p_{10|x_10} \\ p_{00|x_11} & p_{10|x_11} \\ . \\ . \\ . \\ p_{00|x_k0} & p_{10|x_k0} \\ p_{00|x_k1} & p_{10|x_k1} \end{bmatrix} \begin{bmatrix} \frac{1}{p_{0|0}} \\ \frac{1}{p_{0|1}} \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ . \\ . \\ . \\ 1 \\ 1 \end{bmatrix}, \tag{A5}$$

To identify the unknown parameters, the equations should only have one solution. Thus, the rank of coefficient matrix and augmented matrix is 2. Suppose the coefficient matrix and augmented matrix of this equation are $C_1$ and $\bar{C}_1$, respectively, which means:

$$C_1 = \begin{bmatrix} p_{00|x_10} & p_{10|x_10} \\ p_{00|x_11} & p_{10|x_11} \\ . \\ . \\ . \\ p_{00|x_k0} & p_{10|x_k0} \\ p_{00|x_k1} & p_{10|x_k1} \end{bmatrix} \text{ and } \bar{C}_1 = \begin{bmatrix} p_{00|x_10} & p_{10|x_10} & 1 \\ p_{00|x_11} & p_{10|x_11} & 1 \\ . \\ . \\ p_{00|x_k0} & p_{10|x_k0} & 1 \\ p_{00|x_k1} & p_{10|x_k1} & 1 \end{bmatrix}$$

According to the theory of linear equation, when the rank of the coefficient matrix, the rank of the augmented matrix and the number of unknowns are the same ($r(C_1) = r(\bar{C}_1) = 2$),, the equation has a unique solution.

(2): Because $R_Z \perp\!\!\!\perp X|(Y,Z)$, we can spilt the joint distribution into $P(X,Y,Z,R_Z) = P(X,Y)P(Z|X,Y)P(R_Z|Y,Z)$. Divide both sides by $P(X,Y)P(R_Z|X,Y,Z)$ and consider the missing mechanism, and we can get:

$$\frac{P(Z, R_Z=0|X,Y)}{P(R_Z=0|Y,Z)} = P(Z|X,Y),$$

and

$$\sum_{z=0,1} \frac{P(Z=z, R_Z=0|X,Y)}{P(R_Z=0|Y,Z=z)} = \sum_{z=0,1} P(Z=z|X,Y) = 1.$$

$P(Z, R_Z=0|X,Y)$ is identifiable, and we need to prove that $P(Z|X,Y)$ and $P(R_Z=0|Y,Z)$ are also identifiable. Similarly, we can get the following equations:

$$\begin{pmatrix} p_{00|x_10} & p_{10|x_10} \\ \cdots \\ p_{00|x_k0} & p_{10|x_k0} \end{pmatrix} \begin{pmatrix} \frac{1}{p_{0|00}} \\ \frac{1}{p_{0|01}} \end{pmatrix} = \begin{pmatrix} 1 \\ \cdots \\ 1 \end{pmatrix} \tag{A6}$$

$$\begin{pmatrix} p_{00|x_11} & p_{10|x_11} \\ \cdots \\ p_{00|x_k1} & p_{10|x_k1} \end{pmatrix} \begin{pmatrix} \frac{1}{p_{0|10}} \\ \frac{1}{p_{0|11}} \end{pmatrix} = \begin{pmatrix} 1 \\ \cdots \\ 1 \end{pmatrix} \tag{A7}$$

To identify the unknown parameters, the equations should only have one solution. Thus, the rank of coefficient matrices and augmented matrices is 2. Suppose the coefficient matrix and augmented matrix of this equation are $C_{21}$, $C_{22}$, and $\bar{C}_{21}$, $\bar{C}_{22}$ respectively, which means:

$$
C_{21} = \begin{bmatrix} p_{00|x_1 0} & p_{10|x_1 0} \\ \ldots \\ p_{00|x_k 0} & p_{10|x_k 0} \end{bmatrix}, \quad \bar{C}_{21} = \begin{bmatrix} p_{00|x_1 0} & p_{10|x_1 0} & 1 \\ \ldots \\ p_{00|x_k 0} & p_{10|x_k 0} & 1 \end{bmatrix},
$$

$$
C_{22} = \begin{bmatrix} p_{00|x_1 1} & p_{10|x_1 1} \\ \ldots \\ p_{00|x_k 1} & p_{10|x_k 1} \end{bmatrix}, \quad \bar{C}_{22} = \begin{bmatrix} p_{00|x_1 1} & p_{10|x_1 1} & 1 \\ \ldots \\ p_{00|x_k 1} & p_{10|x_k 1} & 1 \end{bmatrix},
$$

According to the theory of linear equations, when the rank of the coefficient matrix, the rank of the augmented matrix, and the number of unknowns are the same($r(C_{21}) = r(\bar{C}_{21}) = r(C_{22}) = r(\bar{C}_{22}) = 2$), the equation has a unique solution.

(3): Because $Y \perp\!\!\!\perp R_Z | (X,Z)$, $P(Y|X,Z) = P(Y|X,Z,R_Z = 1)$ and $P(Y|X,Z,R_Z = 1)$ are, thus, identifiable, and then $P(Y|X,Z)$ is identifiable. In addition, for $P(X)$, similarly, $P(X,Y,Z,R_Z) = P(X,Y)P(Z|X,Y)P(R_Z|X,Y,Z) = P(X,Y)P(Z|X,Y)P(R_Z|X,Z)$. Divide both sides by $P(X,Y)P(R_Z|X,Z)$, and we can get:

$$
\frac{P(Z, R_Z = 0|X,Y)}{P(R_Z = 0|X,Z)} = P(Z|X,Y),
$$

and

$$
\sum_{z=0,1} \frac{P(Z = z, R_Z = 0|X,Y)}{P(R_Z = 0|X,Z = z)} = \sum_{z=0,1} P(Z = z|X,Y) = 1.
$$

To identify $P(X)$ and $P(Y|X,Z)$, we have to identify $P(X,Y,Z,R_Z)$ first. In the equation mentioned above, $P(X,Y)$ is identifiable, so we have to prove that $P(Z|X,Y)$ and $P(R_Z = 0|X,Z)$ are identifiable. Similarly, we can get:

$$
\begin{bmatrix} p_{00|x_1 0} & p_{10|x_1 0} \\ p_{00|x_1 1} & p_{10|x_1 1} \end{bmatrix} \begin{bmatrix} \frac{1}{p_{0|x_1 0}} \\ \frac{1}{p_{0|x_1 1}} \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix},
$$

$$
\ldots
$$

$$
\begin{bmatrix} p_{00|x_k 0} & p_{10|x_k 0} \\ p_{00|x_k 1} & p_{10|x_k 1} \end{bmatrix} \begin{bmatrix} \frac{1}{p_{0|x_k 0}} \\ \frac{1}{p_{0|x_k 1}} \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}.
$$

To identify the unknown parameters, the equations should only have one solution; it is equivalent to:

$$
\begin{bmatrix} p_{00|x_i 0} & p_{10|x_i 0} \\ p_{00|x_i 1} & p_{10|x_i 1} \end{bmatrix} \neq 0 \quad i = 1,2\ldots k,
$$

$$
\Leftrightarrow \frac{p_{00|x_i 0}}{p_{00|x_i 0}} \neq \frac{p_{10|x_i 1}}{p_{10|x_i 1}}
$$

$$
\Leftrightarrow Z \not\perp\!\!\!\perp Y | (R_Z = 0, X = x_i) \quad (i = 1,2\ldots k)
$$

In sum, if the condition $Z \not\perp\!\!\!\perp Y | (R_Z = 0, X = x_i)$ $(i = 1,2\ldots k)$ is satisfied, we can identify the joint distribution $P(X,Y,Z,R_Z)$, and then we can identify the TBR and THR naturally.  □

## References

1. Imbens, G.; Rubin, D. *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction*; Cambridge University Press: Cambridge, UK, 2015.
2. Goetghebeur, E.; Molenberghs, G. Causal inference in a placebo-controlled clinical trial with binary outcome and ordered compliance. *J. Am. Stat. Assoc.* **1996**, *91*, 928–934. [CrossRef]
3. Berger, V.; Rezvani, A.; Makarewicz, V. Direct effect on validity of response run-in selection in clinical trials. *Control. Clin. Trials* **2003**, *24*, 156–166. [CrossRef] [PubMed]
4. Shen, C.; Jeong, J.; Li, X.; Chen, P.S.; Buxton, A. Treatment benefit and treatment harm rate to characterize heterogeneity in treatment effect. *Biometrics* **2013**, *69*, 724–731. [CrossRef] [PubMed]
5. Yin, Y.; Zhou, X.H.; Geng, Z.; Lu, F. Assessing The Heterogeneity Of Treatment Effects By Identifying The Treatment Benefit Rate And Treatment Harm Rate. *Stat. Sin.* **2018**, *28*, 137–156.
6. Shen, C.; Hu, Y.; Li, X.; Wang, Y.; Chen, P.S.; Buxton, A.E. Identification of subpopulations with distinct treatment benefit rate using the Bayesian tree. *Biom. J.* **2016**, *58*, 1357–1375. [CrossRef] [PubMed]
7. Yin, Y.; Liu, L.; Geng, Z. Assessing the treatment effect heterogeneity with a latent variable. *Stat. Sin.* **2018**, *28*, 115–135.
8. Scharfstein, D.; Onicescu, G.; Goodman, S.; Whitaker, R. Analysis of subgroup effects in randomized trials when subgroup membership is missing: Application to the second multicenter automatic defibrillator intervention trial. *J. R. Stat. Soc. Ser. C (Appl. Stat.)* **2011**, *60*, 607–617. [CrossRef]
9. Girard, T.D.; Kress, J.P.; Fuchs, B.D.; Thomason, J.W.; Schweickert, W.D.; Pun, B.T.; Taichman, D.B.; Dunn, J.G.; Pohlman, A.S.; Kinniry, P.A.; et al. Efficacy and safety of a paired sedation and ventilator weaning protocol for mechanically ventilated patients in intensive care (Awakening and Breathing Controlled trial): A randomised controlled trial. *Lancet* **2008**, *371*, 126–134. [CrossRef] [PubMed]
10. Egleston, B.; Wong, Y. Sensitivity analysis to investigate the impact of a missing covariate on survival analyses using cancer registry data. *Stat. Med.* **2009**, *28*, 1498–1511. [CrossRef] [PubMed]
11. Chen, H.; Geng, Z.; Zhou, X.H. Identifiability and estimation of causal effects in randomized trials with noncompliance and. completely nonignorable missing data (with discussion). *Biometrics* **2009**, *65*, 675–682. [CrossRef] [PubMed]
12. Ma, W.; Geng, Z.; Hu, Y. Identification of graphical models for nonignorable nonresponse of binary outcomes in. longitudinal studies. *J. Multivar. Anal.* **2003**, *87*, 24–45. [CrossRef]
13. Frangakis, C.E.; Rubin, D.B. Addressing complications of intention-to-treat analysis in the combined presence of all-or-none treatment-noncompliance and subsequent missing outcomes. *Biometrika* **1999**, *86*, 365–379. [CrossRef]
14. Egleston, B.L.; Scharfstein, D.O.; MacKenzie, E. On estimation of the survivor average causal effect in observational studies when important confounders are missing due to death. *Biometrics* **2009**, *65*, 497–504. [CrossRef] [PubMed]
15. Frangakis, C.E.; Rubin, D.B.; An, M.W.; MacKenzie, E. Principal stratification designs to estimate input data missing due to death (with discussion). *Biometrics* **2007**, *63*, 641–649. [CrossRef] [PubMed]
16. Yan, W.; Hu, Y.; Geng, Z. Identifiability of causal effects for binary variables with baseline data missing due to death. *Biometrics* **2012**, *68*, 121–128. [CrossRef] [PubMed]
17. Rathouz, P.J. Identifiability assumptions for missing covariate data in failure time regression models. *Biostatistics* **2007**, *8*, 345–356. [CrossRef] [PubMed]
18. Little, R.J.; Zhang, N. Subsample ignorable likelihood for regression analysis with missing data. *J. R. Stat. Soc. Ser. C (Appl. Stat.)* **2011**, *60*, 591–605. [CrossRef]
19. Little, R.J.A.; Rubin, D.B. *Statistical Analysis with Missing Data*; John Wiley & Sons: Hoboken, NJ, USA, 2019.
20. Tingley, D.; Yamamoto, T.; Keele, L.; Imai, K. Mediation: R Package for Causal Mediation Analysis. R Package Version 4.2. 2012. Available online: http://CRAN.R-project.org/package=mediation (accessed on 8 October 2023).
21. Ding, P.; Geng, Z. Identifiability of subgroup causal effects in randomized experiments with nonignorable missing covariates. *Stat. Med.* **2014**, *33*, 1121–1133. [CrossRef] [PubMed]