mathematics

MDPI

Article

Non-Parametric Non-Inferiority Assessment in a Three-Arm Trial with Non-Ignorable Missing Data

Wei Li, Yunqi Zhang and Niansheng Tang *

Yunnan Key Laboratory of Statistical Modeling and Data Analysis, Yunnan University, Kunming 650091, China
* Correspondence: nstang@ynu.edu.cn; Tel.: +86-871-65032416

Abstract: A three-arm non-inferiority trial including a placebo is usually utilized to assess the non-inferiority of an experimental treatment to a reference treatment. Existing methods for assessing non-inferiority mainly focus on the fully observed endpoints. However, in some clinical trials, treatment endpoints may be subject to missingness for various reasons, such as the refusal of subjects or their migration. To address this issue, this paper aims to develop a non-parametric approach to assess the non-inferiority of an experimental treatment to a reference treatment in a three-arm trial with non-ignorable missing endpoints. A logistic regression is adopted to specify a non-ignorable missingness data mechanism. A semi-parametric imputation method is proposed to estimate parameters in the considered logistic regression. Inverse probability weighting, augmented inverse probability weighting and non-parametric methods are developed to estimate treatment efficacy for known and unknown parameters in the considered logistic regression. Under some regularity conditions, we show asymptotic normality of the constructed estimators for treatment efficacy. A bootstrap resampling method is presented to estimate asymptotic variances of the estimated treatment efficacy. Three Wald-type statistics are constructed to test the non-inferiority based on the asymptotic properties of the estimated treatment efficacy. Empirical studies show that the proposed Wald-type test procedure is robust to the misspecified missingness data mechanism, and behaves better than the complete-case method in the sense that the type I error rates for the former are closer to the pre-given significance level than those for the latter.

Keywords: bootstrap resampling; imputation; non-inferiority assessment; non-ignorable missing data; three-arm trial

MSC: 62G05; 62D10

Citation: Li, W.; Zhang, Y.; Tang, N. Non-Parametric Non-Inferiority Assessment in a Three-Arm Trial with Non-Ignorable Missing Data. *Mathematics* **2023**, *11*, 246. https://doi.org/10.3390/math11010246

Academic Editor: Leonid V. Bogachev

Received: 25 October 2022
Revised: 27 December 2022
Accepted: 28 December 2022
Published: 3 January 2023

1. Introduction

Non-inferiority (NI) trials are often performed to verify that the efficacy of an experimental treatment with low toxicity or small side-effects is non-inferior to that of a reference treatment by more than a pre-given small margin [1,2]. Many methods have been presented to assess the NI of an experimental treatment to a reference treatment via the efficacy in a two-arm NI trial. For example, see Tang et al. [3] for a score test via relative risk in a matched-pair NI trial; Tang et al. [4] for exact and approximate unconditional confidence intervals for rate difference based on a score test statistic in a small-sample paired NI trial; Wellek [5] for frequentist and Bayesian approaches to testing NI in a matched-pair design with binary endpoints; Freitag et al. [6] for a non-parametric approach to testing NI with censored data; Arboretti et al. [7] and Pesarin et al. [8] for a permutation test in a non-inferiority trial, and Gamalo et al. [9] for a Bayesian method for testing NI with normally distributed endpoints. However, it is widely recognized that there are two key problems for two-arm NI trials [10]. The first issue is the selection of the NI margin (i.e., the clinically acceptable amount or a combination of statistical reasoning and clinical judgement), and the second is the evaluation of assay sensitivity (i.e., the ability of a trial to

distinguish an effective treatment from a less effective or ineffective treatment). To solve the aforementioned problems, if ethically acceptable and practically feasible, a three-arm trial including a placebo, which is called a three-arm NI trial, is usually conducted to assess the NI of an experimental treatment to the active reference treatment [11].

Many methods have been developed to draw statistical inferences based on a three-arm NI trial over the past years. For example, Pigeot et al. [12] studied an NI assessment problem via mean difference in a three-arm trial with normally distributed endpoints; Tang and Tang [10] proposed two asymptotic approaches to testing NI via a rate difference for binary outcomes; Mielke and Munk [13] considered the NI testing problem for Poisson-distributed endpoints; Lui and Chang [14] discussed the NI testing problem via a generalized odds ratio for ordinal data; Brannath et al. [15] considered an NI adaptive testing and sample size determination problem in a three-arm trial with normally distributed endpoints; Tang et al. [16] developed exact and approximate unconditional, and bootstrap-resampling-based approaches to testing NI for binary outcomes; Tang and Yu [17] presented a hybrid approach to constructing simultaneous confidence intervals for simultaneously assessing NI and assay sensitivity for binary endpoints; Tang and Yu [18] utilized two Bayesian approaches (i.e., posterior variance and Bayes factor approaches) to determine the sample size required in a three-arm NI trial with binary endpoints; Paul et al. [19] presented both frequentist and Bayesian procedures for testing NI via the risk difference in a three-arm trial with binary endpoints; Homma and Diamon [20] investigated the assay sensitivity hypothesis and the sample size calculation problem for gold-standard NI trials with two fixed margins and negative binomial endpoints; Ghosh et al. [21] presented a new method to test NI for Poisson-distributed endpoints; Ghosh et al. [22] considered a hierarchical testing procedure with two stages in three-arm NI trials; Scharpenberg and Brannath [23] discussed simultaneous confidence intervals of risk differences in three-arm non-inferiority trials; and Tang and Liang [24] constructed two simultaneous confidence intervals for assessing NI and assay sensitivity in a three-arm trial. However, when misspecifying the distributions of treatment endpoints, statistical inference obtained with the aforementioned methods may be misleading or unreasonable. To this end, a number of non-parametric methods were proposed to make statistical inference on three-arm NI trials under an unknown distribution assumption of endpoints. For example, see Munzel [25] for a rank-based NI test and Tseng and Hsu [26] for binomially distributed outcomes. The aforementioned methods were developed for the fully observed endpoints in a three-arm trial.

However, in some clinical trials, treatment endpoints may be subject to missingness occurring for various reasons, such as unwillingness of some respondents to answer sensitivity questions, loss of information caused by uncontrollable factors, or drop-out from the study in clinical trials [27]. For example, for a clinical trial associated with HIV patients in the AIDS Clinical Trial Group (ACTG) Study 193A, the primary endpoint was the CD4 cell count, which was scheduled to be observed at baseline and eight-week intervals during the follow-up period, potentially subject to missingness due to skipped visits and dropouts. In this study, 1309 patients were randomly assigned to one of the following four daily regimens: zidovudine alternating monthly with 400 mg didanosine (regarded as "Treatment 1"), zidovudine plus 2.25 mg of zalcitabine (regarded as "Treatment 2"), zidovudine plus 400 mg of didanosine (regarded as "Treatment 3"), zidovudine plus 400 mg of didanosine plus 400 mg of nevirapine (regarded as "Treatment 4"). As an illustration, we here take "Treatment 1", "Treatment 2" and "Treatment 4" as the placebo, reference and experiment, respectively, let the log-transformed CD4 cell counts (i.e., log(1 + CD4 cell counts at time interval (4,12])) be the treatment endpoints, and regard log(baseline measurement + 1) as an instrument variable for skipped visits or dropouts. Because the baseline measurements were considered before the treatments were assigned, it was reasonable to assume that the dropouts were missing not at random (MNAR) or due to non-ignorable missing. The average missing proportions of endpoints for the placebo, reference and experiment treatments were 29.74%, 30.99% and 28.65%, respectively. Our

main purpose is to test the NI of treatment 4 to treatment 2 in terms of the assay sensitivity and the internal validity of treatment 4 in a three-arm trial with unknown distributed endpoints in the presence of non-ignorable missing endpoints.

For the above described example, the simplest and most intuitive method for handling missing data is the well-known complete-case ('CC') method, i.e., deleting subjects with missing data. But the CC method may lead to a biased estimator of treatment efficacy when the missingness data mechanism does not involve missing completely at random. To this end, several alternative methods have been proposed to make statistical inferences in two-arm trials with missing endpoints. For example, Choi and Stablein [28] considered the problem of testing the equality of two treatments in a paired two-arm trial with missing at random (MAR) endpoints based on large sample theory, while Tang and Tang [29] developed unconditional exact procedures for testing the equality of two treatments in a paired two-arm trial with MAR endpoints. In addition, some permutation tests were proposed for endpoints with missing data in two-arm trials, for example, see Maritz [30], Yu et al. [31], Pesarin [32], and Pesarin et al. [33]. However, the aforementioned studies mainly focused on equivalence assessment in two-arm trials with a MAR assumption based on two independent binomial distributions for endpoints with non-ignorable missing and a multinomial distribution for the fully observed endpoints. Moreover, to our knowledge, there has been little work undertaken on NI assessment in three-arm trials with unknown distributed endpoints and non-ignorable missing data. Hence, this paper aims to develop a non-parametric approach to testing NI in a three-arm trial with a mixed unknown distribution of endpoints and an MNAR assumption of missing endpoints.

There are many approaches to handling non-ignorable missing data. For example, see Robins et al. [34] for an inverse probability weighting (IPW) method, Lee and Tang [35] and Wang and Tang [36] for Bayesian approaches combining the Gibbs sampler and Metropolis–Hastings algorithm, Kim and Yu [37] for a semi-parametric approach to estimating mean functions in the presence of non-ignorable missing responses, and Tang et al. [38] for an empirical likelihood method for generalized estimating equations with non-ignorable missing data due to certain merits of empirical likelihood, such as feasibly incorporating auxiliary information to improve the efficiency of parameter estimation [39]. Choi and Stablein [40] and Li et al. [41] investigated the equivalence test problem in a paired two-arm trial with non-ignorable missing endpoints under some known distribution assumptions for treatment endpoints. However, the aforementioned approaches cannot be directly used to test NI in a three-arm trial with non-ignorable missing endpoints due to the complexity of the considered test problem, including the imputation of missing endpoints, the estimation problem of treatment efficacy under unknown distribution assumptions of treatment endpoints, and the critical value determination of test statistics at some pre-given significance level.

The main contributions of this paper include: (i) presentation of a logistic regression to specify the propensity score function associated with respondent endpoints; (ii) proposal of IPW, augmented IPW (AIPW) and non-parametric imputation methods to estimate treatment efficacy in the presence of non-ignorable missing endpoints; (iii) development of a semi-parametric imputation method to estimate unknown parameters in the considered logistic regression by imputing mean score functions rather than missing endpoints using a kernel non-parametric regression method; (iv) establishment of some asymptotic properties of the estimated treatment efficacy; (v) refinement of a bootstrap-resampling method to consistently estimate asymptotic variances of the estimated treatment efficacy; (vi) construction of three Wald-type statistics to test the NI of an experimental treatment to a reference treatment in a three-arm trial with unknown distributed and non-ignorable missing endpoints.

The rest of this paper is organized as follows: Section 2 describes a three-arm NI trial with MNAR endpoints. Section 3 discusses the estimation problem of treatment efficacy and propensity score function. The asymptotic properties of the estimated treatment efficacy and the resultant Wald-type statistics for testing NI are given in Section 4. The simulation

studies investigating the finite sample performance of the proposed test statistics are described in Section 5. A real example taken from the ACTG study is illustrated using the proposed method in Section 6. Some concluding remarks are given in Section 7. Technical details are presented in Appendix A.

2. A Three-Arm NI Trail with MNAR Endpoints

2.1. A Three-Arm NI Trial

For a three-arm randomized clinical trial with experimental (E), reference (R) and placebo (P) treatments, we assume that their corresponding clinical endpoints Y_E, Y_R and Y_P independently follow unknown distributions $f_E(y_E|\mu_E)$, $f_R(y_R|\mu_R)$ and $f_P(y_P|\mu_P)$, respectively, where μ_E, μ_R and μ_P are their corresponding treatment efficacies, respectively. Generally, we assume that a larger value of treatment efficacy indicates a more favorable treatment.

Following Hida and Tango [42], to test the NI of the experimental treatment to the reference treatment in terms of assay sensitivity in a three-arm trial, we need to simultaneously demonstrate (i) the superiority of the experimental treatment to placebo, (ii) the NI of the experimental treatment to the reference treatment for a pre-specified maximal clinically irrelevant or NI margin $\delta > 0$, and (iii) the superiority of the reference treatment to placebo by more than δ. That is, μ_E, μ_R and μ_P must satisfy the following inequalities: $\mu_P < \mu_R - \delta < \mu_E$, which leads to consideration of the following hypothesis-testing problem:

$$\begin{array}{l} H_0 : \mu_E \le \mu_R - \delta \quad \text{versus} \quad H_1 : \mu_E > \mu_R - \delta, \\ K_0 : \mu_R \le \mu_P + \delta \quad \text{versus} \quad K_1 : \mu_R > \mu_P + \delta. \end{array} \tag{1}$$

Clearly, simultaneously rejecting H_0 and K_0 at some pre-given significance level yields the above desirable inequalities: $\mu_P < \mu_R - \delta < \mu_E$ indicating the NI of the experimental treatment to the reference treatment and assay sensitivity. Generally, the selection of the NI margin δ should combine statistical reasoning and clinical judgement [17]. In a similar way to many three-arm trial studies, the fraction margin approach can be used to specify δ.

Following Kieser and Friede [43], δ can be mathematically expressed as a positive fraction f of the unknown efficacy difference between the reference treatment and placebo, i.e., $\delta = g(\mu_R - \mu_P)$, where g lies in the interval $[0, 1]$. The NI margin δ defined above indicates that the condition of assay sensitivity holds, i.e., $\mu_R - \mu_P > 0$. Following the argument of Ghosh et al. [44], one can take $g = 1/2$ or $1/3$. To explain the hypotheses considered above, we set $a = 1 - g \in (0, 1)$, whose different values have different statistical meanings [12]. Under the above assumption, we only need to test H_0 rather than hypothesis (1). That is, for the NI margin δ defined above, we only need to test the following hypothesis:

$$H_0 : \mu_E - a\mu_R - (1-a)\mu_P \le 0 \ \text{ versus } \ H_1 : \mu_E - a\mu_R - (1-a)\mu_P > 0. \tag{2}$$

Rejecting H_0 at some pre-given significance level indicates the NI of the experimental treatment to the reference treatment under the condition of assay sensitivity. For simplicity, we denote $\psi(\mu) = \mu_E - a\mu_R - (1-a)\mu_P$, where $\mu = \{\mu_E, \mu_R, \mu_P\}$. In this case, the hypothesis (2) can re-expressed as

$$\tilde{H}_0 : \psi(\mu) \le 0 \quad \text{versus} \quad \tilde{H}_1 : \psi(\mu) > 0. \tag{3}$$

2.2. Missingness Data Mechanism

Let $\{Y_{\ell i} : i = 1, \ldots, n_\ell\}$ be the clinical observations of Y_ℓ for n_ℓ subjects randomly assigned to treatment ℓ for $\ell = E, R, P$. Here, we assume that $Y_{\ell i}$'s may be subject to missingness, let $D_{\ell i}$ be the indicator of non-missing observation $Y_{\ell i}$, i.e., $D_{\ell i} = 1$ if $Y_{\ell i}$ is observed, and $D_{\ell i} = 0$ if $Y_{\ell i}$ is missing, and define $\boldsymbol{X}_{\ell i}$ as a vector of covariates for $\ell = E, R, P$ and $i = 1, \ldots, n_\ell$. It is also assumed that $\boldsymbol{X}_{\ell i}$'s are fully observed, $D_{\ell i_1}$ is independent of $D_{\ell i_2}$ for $i_1 \ne i_2 \in \{1, \ldots, n_\ell\}$, and $D_{\ell i}$ depends on the observed covariates $\boldsymbol{X}_{\ell i}$ and missing observation $Y_{\ell i}$, which indicates that the considered non-missingness data

mechanism is non-ignorable. Under the above assumption, we consider the following non-missingness data mechanism model:

$$\pi_{\ell i}(\boldsymbol{\eta}_\ell; \boldsymbol{X}_{\ell i}, Y_{\ell i}) = \Pr(D_{\ell i} = 1 | \boldsymbol{X}_{\ell i}, Y_{\ell i}; \boldsymbol{\eta}_\ell),\ \ell = E, R, P,\ \ i = 1, \ldots, n_\ell,$$

where $\boldsymbol{\eta}_\ell$ is a vector of unknown parameters to be estimated, and $\pi_{\ell i}(\boldsymbol{\eta}_\ell; \boldsymbol{X}_{\ell i}, Y_{\ell i})$ is usually called the propensity score function in the missing data literature.

Many methods can be employed to specify the propensity score function $\pi_{\ell i}(\boldsymbol{\eta}_\ell; \boldsymbol{X}_{\ell i}, Y_{\ell i})$. For example, see Lee and Tang [35] for a logistic regression, Kim and Yu [37] and Tang et al. [38] for an exponential tilting model, and Wang and Tang [36] for a probit regression model. Here, similarly to Lee and Tang [35], we consider the following logistic regression model for $\pi_{\ell i}(\boldsymbol{\eta}_\ell; \boldsymbol{X}_{\ell i}, Y_{\ell i})$:

$$\text{logit}\{\pi_{\ell i}(\boldsymbol{\eta}_\ell; \boldsymbol{X}_{\ell i}, Y_{\ell i})\} = \alpha_{\ell 0} + \boldsymbol{\alpha}_{\ell 1}^\top \boldsymbol{X}_{\ell i} + \gamma_\ell Y_{\ell i},\ \ \ell = E, R, P, i = 1, \ldots, n_\ell,$$

where $\text{logit}(c) = \log\{c/(1-c)\}$, and $\boldsymbol{\eta}_\ell = (\alpha_{\ell 0}, \boldsymbol{\alpha}_{\ell 1}^\top, \gamma_\ell)^\top$. It is well-known that, when γ_ℓ is unknown, the above specified logistic regression model is unidentifiable. To address this issue, we decompose $\boldsymbol{X}_{\ell i}$ as $\boldsymbol{X}_{\ell i} = (\boldsymbol{Z}_{\ell i}^\top, \boldsymbol{U}_{\ell i}^\top)^\top$, where $\boldsymbol{Z}_{\ell i}$ may be associated with the propensity score function, and $\boldsymbol{U}_{\ell i}$ is a vector of instrumental variables that is not directly associated with the propensity score function but related to observations $Y_{\ell i}$. In this case, we can consider the following propensity score function

$$\text{logit}\{\pi_{\ell i}(\boldsymbol{\eta}_\ell; \boldsymbol{Z}_{\ell i}, Y_{\ell i})\} = \alpha_{\ell 0} + \boldsymbol{\alpha}_{\ell 1}^\top \boldsymbol{Z}_{\ell i} + \gamma_\ell Y_{\ell i},\ \ \ell = E, R, P, i = 1, \ldots, n_\ell. \tag{4}$$

Clearly, when $\gamma_\ell = 0$, the above defined missingness data mechanism reduces to MAR.

3. Estimation of Treatment Efficacy

3.1. Estimating Treatment Efficacy

When the endpoints are completely observed, treatment efficacy μ_ℓ can be consistently estimated by its corresponding sample mean, i.e., $\hat{\mu}_\ell = n_\ell^{-1} \sum_{i=1}^{n_\ell} Y_{\ell i}$ for $\ell = E, R, P$.

When $Y_{\ell i}$'s are subject to missingness and the true propensity score function $\pi_{\ell i}(\boldsymbol{\eta}_\ell; \boldsymbol{Z}_{\ell i}, Y_{\ell i})$ is known, the IPW method can be employed to estimate μ_ℓ for $\ell = E, R, P$. That is, μ_ℓ can be estimated by

$$\hat{\mu}_\ell^{HT} = \frac{1}{n_\ell} \sum_{i=1}^{n_\ell} \frac{D_{\ell i}}{\pi_{\ell i}(\boldsymbol{\eta}_\ell; \boldsymbol{Z}_{\ell i}, Y_{\ell i})} Y_{\ell i},\ \ \ell = E, R, P. \tag{5}$$

Note that the above defined estimator $\hat{\mu}_\ell^{HT}$ may be sensitive to the misspecification of the propensity score function. To address this issue, an imputation technique is adopted to construct a consistent estimator of μ_ℓ in the presence of MNAR. That is, let $m_{\ell i}^0(\gamma_\ell; \boldsymbol{Z}_{\ell i}) = E(Y_{\ell i} | \boldsymbol{Z}_{\ell i}, D_{\ell i} = 0)$, an imputation-based estimator of μ_ℓ has the form

$$\hat{\mu}_\ell^{RI} = \frac{1}{n_\ell} \sum_{i=1}^{n_\ell} \Big\{ D_{\ell i} Y_{\ell i} + (1 - D_{\ell i}) m_{\ell i}^0(\gamma_\ell; \boldsymbol{Z}_{\ell i}) \Big\},\ \ \ell = E, R, P. \tag{6}$$

The AIPW approach can also be utilized to estimate μ_ℓ in the presence of MNAR. That is, an AIPW-based estimator of μ_ℓ can be expressed as

$$\hat{\mu}_\ell^{AI} = \frac{1}{n_\ell} \sum_{i=1}^{n_\ell} \left\{ \frac{D_{\ell i}}{\pi_{\ell i}(\boldsymbol{\eta}_\ell; \boldsymbol{Z}_{\ell i}, Y_{\ell i})} Y_{\ell i} + \left(1 - \frac{D_{\ell i}}{\pi_{\ell i}(\boldsymbol{\eta}_\ell; \boldsymbol{Z}_{\ell i}, Y_{\ell i})} \right) m_{\ell i}^0(\gamma_\ell; \boldsymbol{Z}_{\ell i}) \right\},\ \ \ell = E, R, P. \tag{7}$$

In many clinical trials, the cumulative distribution functions of $Y_{\ell i}$'s are usually unknown; thus, $m_{\ell i}^0(\gamma_\ell; \boldsymbol{Z}_{\ell i})$'s are also unknown. On the other hand, $\pi_{\ell i}(\boldsymbol{\eta}_\ell; \boldsymbol{Z}_{\ell i}, Y_{\ell i})$ is also unknown in the presence of MNAR. Hence, it is impossible to directly evaluate $\hat{\mu}_\ell^{HT}$, $\hat{\mu}_\ell^{RI}$

and $\hat{\mu}_\ell^{AI}$ using the above defined forms. In what follows, we consider the estimation problem of $m_{\ell i}^0(\gamma_\ell; \mathbf{Z}_{\ell i})$ and $\pi_{\ell i}(\boldsymbol{\eta}_\ell; \mathbf{Z}_{\ell i}, Y_{\ell i})$.

3.2. Estimation of Conditional Mean $m_{\ell i}^0(\gamma_\ell; \mathbf{Z}_{\ell i})$

Here a non-parametric method given in Tang et al. [38] is adopted to estimate $m_{\ell i}^0(\gamma_\ell; \mathbf{Z}_{\ell i})$ in the presence of MNAR.

Let $f_{0\ell}(y_{\ell i})$ and $f_{1\ell}(y_{\ell i})$ be the conditional probability densities of $Y_{\ell i}$ given $D_{\ell i} = 0$ and $D_{\ell i} = 1$, respectively. Following the argument of Tang et al. [38], we have

$$f_{0\ell}(y_{\ell i}) = f_{1\ell}(y_{\ell i}) \times \frac{O(\boldsymbol{\eta}_\ell; \mathbf{Z}_{\ell i}, Y_{\ell i})}{E\{O(\boldsymbol{\eta}_\ell; \mathbf{Z}_{\ell i}, y_{\ell i}) | \mathbf{Z}_{\ell i}, D_{\ell i} = 1\}}, \quad \ell = E, R, P, \tag{8}$$

where $O(\boldsymbol{\eta}_\ell; \mathbf{Z}_{\ell i}, y_{\ell i}) = \{1 - \pi_{\ell i}(\boldsymbol{\eta}_\ell; \mathbf{Z}_{\ell i}, Y_{\ell i})\} / \pi_{\ell i}(\boldsymbol{\eta}_\ell; \mathbf{Z}_{\ell i}, Y_{\ell i})$. Substituting $\pi_{\ell i}(\boldsymbol{\eta}_\ell; \mathbf{Z}_{\ell i}, Y_{\ell i})$ defined in Equation (4) into (8) leads to

$$f_{0\ell}(y_{\ell i}) = f_{1\ell}(y_{\ell i}) \times \frac{\exp(-\gamma_\ell Y_{\ell i})}{E\{\exp(-\gamma_\ell Y_{\ell i}) | \mathbf{Z}_{\ell i}, D_{\ell i} = 1\}}, \quad \ell = E, R, P, \tag{9}$$

which shows that we can utilize the conditional distribution $f_{1\ell}(y_{\ell i})$ of the observed endpoints rather than that of missing endpoints (i.e., $f_{0\ell}(y_{\ell i})$) to make statistical inferences, where $E(\cdot)$ represents the expectation taken with respect to $f_{1\ell}(y_{\ell i})$. Clearly, when $\gamma_\ell = 0$, we obtain $f_{0\ell}(y_{\ell i}) = f_{1\ell}(y_{\ell i})$.

Following the argument of Tang et al. [38], it follows from Equation (9) that

$$m_{\ell i}^0(\gamma_\ell; \mathbf{Z}_{\ell i}) = \frac{E\{D_{\ell i} Y_{\ell i} \exp(-\gamma_\ell Y_{\ell i}) | \mathbf{Z}_{\ell i}\}}{E\{D_{\ell i} \exp(-\gamma_\ell Y_{\ell i}) | \mathbf{Z}_{\ell i}\}},$$

which implies that a non-parametric regression estimator of $m_{\ell i}^0(\gamma_\ell; \mathbf{Z}_{\ell i})$ can be expressed as

$$\hat{m}_{\ell i}^0(\gamma_\ell; \mathbf{Z}_{\ell i}) = \sum_{k=1}^{n_\ell} \omega_{\ell k0}^i(\gamma_\ell; \mathbf{Z}_{\ell i}) Y_{\ell k}, \tag{10}$$

where $\omega_{\ell k0}^i(\gamma_\ell; \mathbf{Z}_{\ell i})$'s are the weights assigned to $Y_{\ell k}$, and have the form

$$\omega_{\ell k0}^i(\gamma_\ell; \mathbf{Z}_{\ell i}) = \frac{D_{\ell k} \exp(-\gamma_\ell Y_{\ell k}) K_{h_\ell}(\mathbf{Z}_{\ell i} - \mathbf{Z}_{\ell k})}{\sum_{j=1}^{n_\ell} D_{\ell j} \exp(-\gamma_\ell Y_{\ell j}) K_{h_\ell}(\mathbf{Z}_{\ell i} - \mathbf{Z}_{\ell j})}$$

in which $K_{h_\ell}(v) = h_\ell^{-1} K(v/h_\ell)$, $K(\cdot)$ is the multi-dimensional kernel function, and $h_\ell = h_{n_\ell}$ is the bandwidth.

3.3. Estimation of Propensity Score Function

Note that the above considered propensity score function has a parametric form indexed by the parameter vector $\boldsymbol{\eta}_\ell$, which indicates that, if we can obtain the estimation of $\boldsymbol{\eta}_\ell$ (denoted as $\hat{\boldsymbol{\eta}}_\ell$), the estimation of the propensity score function is easily evaluated by $\hat{\pi}_{\ell i}(\hat{\boldsymbol{\eta}}_\ell; \mathbf{Z}_{\ell i}, Y_{\ell i})$. In the following, we discuss the estimation problem of $\boldsymbol{\eta}_\ell$.

The mean score approach of Morikawa et al. [45] is employed here to estimate $\boldsymbol{\eta}_\ell$ based on the observed data $\mathcal{D}_\ell = \{(\boldsymbol{X}_{\ell i}, Y_{\ell i}, D_{\ell i}) : i = 1, \ldots, n_\ell\}$ for $\ell = E, R, P$. For simplicity, we denote $\mathcal{D}_{\text{obs}}^\ell = \{\boldsymbol{X}_\ell, Y_{\text{obs}}^\ell, D_\ell\}$, where $\boldsymbol{X}_\ell = \{\boldsymbol{X}_{\ell i} : i = 1, \ldots, n_\ell\}$, $D_\ell = \{D_{\ell i} : i = 1, \ldots, n_\ell\}$ and Y_{obs}^ℓ is the observed dataset of $Y_{\ell i}$'s.

When the density function $f_\ell(Y_{\ell i} | \mu_\ell; \boldsymbol{X}_{\ell i})$ of $Y_{\ell i}$ is known, the maximum likelihood estimator (MLE) of $\boldsymbol{\eta}_\ell$ can be obtained by maximizing the following likelihood of the observed data $\mathcal{D}_{\text{obs}}^\ell$:

$$\mathcal{L}_{\text{obs}}(\boldsymbol{\eta}_\ell|\mathcal{D}^\ell_{\text{obs}}) = \prod_{i=1}^{n_\ell}\Big[\pi_{\ell i}(\boldsymbol{\eta}_\ell;\mathbf{Z}_{\ell i},Y_{\ell i})f_\ell(Y_{\ell i}|\mu_\ell;\mathbf{X}_{\ell i})\Big]^{D_{\ell i}} \times\Big[\int\{1-\pi_{\ell i}(\boldsymbol{\eta}_\ell;\mathbf{Z}_{\ell i},Y_{\ell i})\}f_\ell(Y_{\ell i}|\mu_\ell;\mathbf{X}_{\ell i})dY_{\ell i}\Big]^{1-D_{\ell i}}.$$

It follows from Morikawa et al. [45] and the mean score theorem that the MLE of $\boldsymbol{\eta}_\ell$ can be obtained by solving the following "mean score equation":

$$\frac{1}{n}\sum_{i=1}^{n_\ell}[D_{\ell i}s(\boldsymbol{\eta}_\ell;D_{\ell i},\mathbf{Z}_{\ell i},y_{\ell i})+(1-D_{\ell i})E\{s(\boldsymbol{\eta}_\ell;D_{\ell i},\mathbf{Z}_{\ell i},Y_{\ell i})|\mathbf{Z}_{\ell i},D_{\ell i}=0\}]=0,$$

where $s(\boldsymbol{\eta}_\ell;D_{\ell i},\mathbf{Z}_{\ell i},Y_{\ell i})$ has the form

$$\begin{aligned}s(\boldsymbol{\eta}_\ell;D_{\ell i},\mathbf{Z}_{\ell i},Y_{\ell i}) &= \frac{\partial}{\partial\boldsymbol{\eta}}\log\Big[\pi_{\ell i}(\boldsymbol{\eta}_\ell;\mathbf{Z}_{\ell i},Y_{\ell i})^{D_{\ell i}}\{1-\pi_{\ell i}(\boldsymbol{\eta}_\ell;\mathbf{Z}_{\ell i},Y_{\ell i})\}^{1-D_{\ell i}}\Big]\\ &= \frac{D_{\ell i}-\pi_{\ell i}(\boldsymbol{\eta}_\ell;\mathbf{Z}_{\ell i},Y_{\ell i})}{\pi_{\ell i}(\boldsymbol{\eta}_\ell;\mathbf{Z}_{\ell i},Y_{\ell i})\{1-\pi_{\ell i}(\boldsymbol{\eta}_\ell;\mathbf{Z}_{\ell i},Y_{\ell i})\}}\dot{\pi}_{\ell i}(\boldsymbol{\eta}_\ell;\mathbf{Z}_{\ell i},Y_{\ell i}),\end{aligned}$$

and $\dot{\pi}_{\ell i}(\boldsymbol{\eta}_\ell;\mathbf{Z}_{\ell i},Y_{\ell i})=\partial\pi_{\ell i}(\boldsymbol{\eta}_\ell;\mathbf{Z}_{\ell i},Y_{\ell i})/\partial\boldsymbol{\eta}_\ell$.

Denote $s_{\ell 0}(\boldsymbol{\eta}_\ell) = E\{s(\boldsymbol{\eta}_\ell;D_{\ell i},\mathbf{Z}_{\ell i},Y_{\ell i})|\mathbf{Z}_{\ell i},D_{\ell i}=0\}$. Again, it follows from Tang et al. [38] and Equation (9) that a non-parametric estimator of $s_{\ell 0}(\boldsymbol{\eta}_\ell)$ is given as

$$\hat{s}_{\ell 0}(\boldsymbol{\eta}_\ell;D_{\ell i},\mathbf{Z}_{\ell i},Y_{\ell i})=\sum_{k=1}^{n_\ell}\omega^i_{\ell k0}(\gamma_\ell;\mathbf{Z}_{\ell i})s(\boldsymbol{\eta}_\ell;D_{\ell k},\mathbf{Z}_{\ell k},Y_{\ell k}),$$

where $\omega^i_{\ell k0}(\gamma_\ell;\mathbf{Z}_{\ell i})$ is defined in Equation (10). Thus, the estimated "mean score equation" can be written as

$$\frac{1}{n}\sum_{i=1}^{n_\ell}\{D_{\ell i}s(\boldsymbol{\eta}_\ell;D_{\ell i},\mathbf{Z}_{\ell i},y_{\ell i})+(1-D_{\ell i})\hat{s}_{\ell 0}(\boldsymbol{\eta}_\ell;D_{\ell i},\mathbf{Z}_{\ell i},Y_{\ell i})\}=0, \tag{11}$$

which shows that the MLE $\hat{\boldsymbol{\eta}}_\ell$ of $\boldsymbol{\eta}_\ell$ can be obtained by solving the non-linear equation (11) with respect to $\boldsymbol{\eta}$.

Once we obtain MLE $\hat{\boldsymbol{\eta}}_\ell$ of $\boldsymbol{\eta}_\ell$, substituting $\hat{\boldsymbol{\eta}}_\ell$ into Equations (4) and (10) leads to the estimated propensity score function $\hat{\pi}_{\ell i}(\hat{\boldsymbol{\eta}}_\ell;\mathbf{Z}_{\ell i},Y_{\ell i})$ and the estimated mean functions $\hat{m}^0_{\ell i}(\hat{\gamma}_\ell,\mathbf{Z}_{\ell i})$. Thus, substituting $\hat{\pi}_{\ell i}(\hat{\boldsymbol{\eta}}_\ell;\mathbf{Z}_{\ell i},Y_{\ell i})$ and $\hat{m}^0_{\ell i}(\hat{\gamma},\mathbf{Z}_{\ell i})$ into Equations (5)–(7) yields non-parametric estimators of treatment efficacy μ_ℓ for $\ell=E,R,P$.

3.4. Dimension Reduction

In some clinical trials, the number of covariates $\mathbf{Z}_\ell\in\mathcal{R}^{d_\ell}$ may be large. In this case, the kernel-based estimators of $s(\boldsymbol{\eta}_\ell;D_{\ell i},\mathbf{Z}_{\ell i},Y_{\ell i})$ and $m^0_{\ell i}(\gamma_\ell,\mathbf{Z}_{\ell i})$ may suffer from the well-known curse of dimensionality. The dimension reduction technique of Tang et al. [38] is used to solve this problem.

Let $\mathcal{G}_\ell:\mathcal{R}^{d_\ell}\to\mathcal{R}$ be a mapping function such that $\mathcal{G}_{\ell i}=\mathcal{G}_\ell(\mathbf{Z}_{\ell i})$ is univariate. In particular, we assume that $E\{s(\boldsymbol{\eta}_\ell;D_{\ell i},\mathbf{Z}_{\ell i},Y_{\ell i})|\mathcal{G}_{\ell i},D_{\ell i}=0\}$ and $E(Y_{\ell i}|\mathcal{G}_{\ell i},D_{\ell i}=0)$ have the same structures as $s_{\ell 0}(\boldsymbol{\eta}_\ell)=E\{s(\boldsymbol{\eta}_\ell;D_{\ell i},\mathbf{Z}_{\ell i},Y_{\ell i})|\mathbf{Z}_{\ell i},D_{\ell i}=0\}$ and $m^0_{\ell i}(\gamma_\ell;\mathbf{Z}_{\ell i})=E(Y_{\ell i}|\mathbf{Z}_{\ell i},D_{\ell i}=0)$, except that $\mathbf{Z}_{\ell i}$ is replaced by $\mathcal{G}_{\ell i}$. Given the MLE $\hat{\boldsymbol{\eta}}_\ell$ of $\boldsymbol{\eta}_\ell$ obtained with the above introduced approach, we can obtain non-parametric dimension reduction estimators of treatment efficacy μ_ℓ for $\ell=E,R,P$.

4. Asymptotic Properties and Test Statistics

4.1. Asymptotic Properties

In the following, we investigate the consistency and asymptotic normality of the proposed estimators $\hat{\mu}^{HT}_\ell$, $\hat{\mu}^{RI}_\ell$, $\hat{\mu}^{AI}_\ell$ with the known and estimated values of parameters $\boldsymbol{\eta}_\ell$.

The notation $\xrightarrow{\mathcal{L}}$ represents convergence in distribution and $\mathcal{N}(\cdot,\cdot)$ denotes the normal distribution.

From Morikawa et al. [45], we obtain the following proposition.

Proposition 1. *Suppose that Assumptions A1–A3 given in the Appendix A hold. The MLE $\hat{\boldsymbol{\eta}}_\ell$ of $\boldsymbol{\eta}_\ell$ satisfies*

$$\sqrt{n_\ell}\left(\hat{\boldsymbol{\eta}}_\ell - \boldsymbol{\eta}_\ell^0\right) \xrightarrow{\mathcal{L}} \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_{\eta\ell}) \text{ as } n_\ell \to \infty,$$

where $\boldsymbol{\eta}_\ell^0$ is the true value of $\boldsymbol{\eta}_\ell$, $\boldsymbol{\Sigma}_{\eta\ell} = \mathcal{I}_{22\ell}^{-1} E\{\exp(-\gamma_\ell^0 Y_{\ell i}) s_{\ell 0}(\boldsymbol{\eta}_\ell^0) s_{\ell 0}(\boldsymbol{\eta}_\ell^0)^\top\} \mathcal{I}_{22\ell}^{-\top}$, $s_{\ell 0}(\boldsymbol{\eta}_\ell^0)$ represents $s_{\ell 0}(\boldsymbol{\eta}_\ell)$ evaluated at $\boldsymbol{\eta}_\ell = \boldsymbol{\eta}_\ell^0$, $\mathcal{I}_{22\ell} = -E\{s_{\ell 0}(\boldsymbol{\eta}_\ell^0)\dot{\pi}_{\ell i}^\top(\boldsymbol{\eta}_\ell^0)/\pi_{\ell i}(\boldsymbol{\eta}_\ell^0)\}$, and $\dot{\pi}_{\ell i}(\boldsymbol{\eta}_\ell) = \dot{\pi}_{\ell i}(\boldsymbol{\eta}_\ell; \mathbf{Z}_{\ell i}, Y_{\ell i})$ and $\pi_{\ell i}(\boldsymbol{\eta}_\ell^0) = \pi_{\ell i}(\boldsymbol{\eta}_\ell^0; \mathbf{Z}_{\ell i}, Y_{\ell i}))$ for $\ell = E, R, P$.

Proof of Proposition 1 can be found in Morikawa et al. [45]. To save space, we omit it. Proposition 1 shows that the MLE $\hat{\boldsymbol{\eta}}_\ell$ of $\boldsymbol{\eta}_\ell$ is consistent and asymptotically distributed as the multivariate normal distribution.

Theorem 1. *Suppose that Assumptions A1–A3 given in the Appendix A hold. For a known value $\boldsymbol{\eta}_\ell^0$ of $\boldsymbol{\eta}_\ell$, given the true value μ_ℓ^0 of μ_ℓ, the proposed estimators $\hat{\mu}_\ell^{HT}$, $\hat{\mu}_\ell^{RI}$ and $\hat{\mu}_\ell^{AI}$ satisfy*

$$\sqrt{n_\ell}(\hat{\mu}_\ell^h - \mu_\ell^0) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma_\ell^2) \quad \text{as} \quad n_\ell \to \infty$$

for $h = HT, RI, AI$ and $\ell = E, R, P$, where $\sigma_\ell^2 = \mathrm{var}(\tau_{\ell i})$ with $\tau_{\ell i} = m_{\ell i}^0(\gamma_\ell; \mathbf{Z}_{\ell i}) + D_{\ell i}\pi_{\ell i}^{-1}(\boldsymbol{\eta}_\ell^0)\{Y_{\ell i} - m_{\ell i}^0(\gamma_\ell; \mathbf{Z}_{\ell i})\}$. In addition, σ_ℓ^2 can be rewritten as $\sigma_\ell^2 = \mathrm{var}(Y_{\ell i}) + E\{[\pi_{\ell i}^{-1}(\boldsymbol{\eta}_\ell^0) - 1][Y_{\ell i} - m_{\ell i}^0(\gamma_\ell; \mathbf{Z}_{\ell i})]^2\}$.

Theorem 1 shows that the proposed estimators of μ_ℓ are consistent and asymptotically distributed as the normal distribution with zero mean and the same variance.

Following the argument of Kim and Yu [37], σ_ℓ^2 can be consistently estimated by

$$\hat{\sigma}_\ell^2 = \frac{1}{n_\ell}\sum_{i=1}^{n_\ell} \hat{\tau}_{\ell i}^2 - \left(\frac{1}{n_\ell}\sum_{i=1}^{n_\ell} \hat{\tau}_{\ell i}\right)^2,$$

where $\hat{\tau}_{\ell i} = \hat{m}_{\ell i}^0(\gamma_\ell^0; \mathbf{Z}_{\ell i}) + D_{\ell i}\hat{\pi}_{\ell i}^{-1}(\boldsymbol{\eta}_\ell^0)\{Y_{\ell i} - \hat{m}_{\ell i}^0(\gamma_\ell^0; \mathbf{Z}_{\ell i})\}$.

When $\boldsymbol{\eta}_\ell$ is unknown, we replace $\boldsymbol{\eta}_\ell$ or γ_ℓ in Equations (5)–(7) by their corresponding consistent estimators $\hat{\boldsymbol{\eta}}_\ell$ or $\hat{\gamma}_\ell$, respectively. Thus, we can obtain their corresponding plug-in estimators (denoted as $\hat{\mu}_\ell^{SHT}$, $\hat{\mu}_\ell^{SRI}$ and $\hat{\mu}_\ell^{SAI}$, respectively) of μ_ℓ.

Theorem 2. *Suppose that Assumptions A1–A3 given in the Appendix A hold, the propensity score function (4) is correctly specified, and Proposition 1 holds. The plug-in estimators $\hat{\mu}_\ell^{SHT}$, $\hat{\mu}_\ell^{SRI}$ and $\hat{\mu}_\ell^{SAI}$ of μ_ℓ satisfy*

$$\sqrt{n_\ell}\left(\hat{\mu}_\ell^h - \mu_\ell^0\right) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma_{\ell,h}^2) \quad \text{as} \quad n_\ell \to \infty\,, \quad h = SHT, SRI, SAI, \quad \ell = E, R, P,$$

where $\sigma_{\ell,h}^2 = \mathrm{var}(e_{\ell,hi})$ with $e_{\ell,hi} = \{D_{\ell i}\pi_{\ell i}^{-1}(\boldsymbol{\eta}_\ell)\{Y_{\ell i} - m_{\ell i}^0(\gamma_\ell; \mathbf{Z}_{\ell i})\} + m_{\ell i}^0(\gamma_\ell; \mathbf{Z}_{\ell i}) - \mu_\ell^0 + \mathcal{I}_{22\ell}^{-1} s_{\ell i}(\boldsymbol{\eta}_\ell) H_{\ell,h}\}$, $s_{\ell i}(\boldsymbol{\eta}_\ell)$ is the ith term in Equation (11), $H_{\ell,SHT} = E\{(\pi_{\ell i}(\boldsymbol{\eta}_\ell) - 1)Y_{\ell i}(1, \mathbf{Z}_{\ell i}^\top, Y_\ell)^\top\}$, $H_{\ell,SRI} = E\{(1 - D_{\ell i})(0, \mathbf{0}_{p_\ell - 1}^\top, (Y_{\ell i} - m_{\ell i}^0(\gamma_\ell; \mathbf{Z}_{\ell i}))^2)^\top\}$, $H_{\ell,SAI} = H_{\ell,SHT} + M_{\ell,SAI}$, $M_{\ell,SAI} = E\{(1 - \pi_{\ell i}(\boldsymbol{\eta}_\ell))m_{\ell i}^0(\gamma_\ell; \mathbf{Z}_{\ell i})(1, \mathbf{Z}_{\ell i}^\top, Y_{\ell i})^\top\}$, $\mathbf{0}_{p_\ell - 1}$ is a $(p_\ell - 1) \times 1$ zero vector and p_ℓ is the number of covariate vector $\mathbf{Z}_{\ell i}$.

Note that the asymptotic variance $\sigma_{\ell,h}^2$ has a complicated form; thus, it is rather difficult to compute the estimate of $\sigma_{\ell,h}^2$. To overcome this difficulty, we utilize a bootstrap-resampling method or empirical jack-knife method to evaluate the estimated asymptotic variances.

4.2. Wald-Type Statistics for Testing $\tilde{H}_0$

In what follows, we construct three Wald-type statistics for testing hypothesis $H_0 : \psi(\boldsymbol{\mu}) \leq 0$ based on the asymptotic properties of three different estimators given in Theorem 2.

Based on the properties of estimator $\hat{\boldsymbol{\mu}}^h = (\hat{\mu}_E^h, \hat{\mu}_R^h, \hat{\mu}_P^h)^\top$ for $\boldsymbol{\mu} = (\mu_E, \mu_R, \mu_P)^\top$, we obtain that (i) $\hat{\psi}(\hat{\boldsymbol{\mu}}^h) = \hat{\mu}_E^h - a\hat{\mu}_R^h - (1-a)\hat{\mu}_P^h$ is a consistent estimator of $\psi(\boldsymbol{\mu})$, (ii) variance of $\hat{\psi}(\hat{\boldsymbol{\mu}}^h)$ is $\text{var}\{\hat{\psi}(\hat{\boldsymbol{\mu}}^h)\} = \text{var}(\hat{\mu}_E^h) + a^2\text{var}(\hat{\mu}_R^h) + (1-a)^2\text{var}(\hat{\mu}_P^h)$, which can consistently be estimated by $\widehat{\text{var}}\{\hat{\psi}(\hat{\boldsymbol{\mu}}^h)\} = \tilde{\sigma}_{E,h}^2/n_E + a^2\tilde{\sigma}_{R,h}^2/n_R + (1-a)^2\tilde{\sigma}_{P,h}^2/n_P$, where $\tilde{\sigma}_{E,h}^2$, $\tilde{\sigma}_{R,h}^2$ and $\tilde{\sigma}_{P,h}^2$ defined in Theorem 2 are the consistent estimators of $\sigma_{E,h}^2$, $\sigma_{R,h}^2$ and $\sigma_{P,h}^2$, respectively, for $h = SHT, SRI, SAI$; (iii) $(\hat{\psi}(\hat{\boldsymbol{\mu}}^h) - \psi(\boldsymbol{\mu}))/\sqrt{\widehat{\text{var}}\{\hat{\psi}(\hat{\boldsymbol{\mu}}^h)\}} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1)$ as $\min\{n_E, n_R, n_P\} \to \infty$. Thus, the Wald-type statistic for testing $\tilde{H}_0 : \psi(\boldsymbol{\mu}) \leq 0$ can be expressed as

$$T_W^h = \frac{\hat{\psi}(\hat{\boldsymbol{\mu}}^h)}{\sqrt{\widehat{\text{var}}\{\hat{\psi}(\hat{\boldsymbol{\mu}}^h)\}}} = \frac{\hat{\mu}_E^h - a\hat{\mu}_R^h - (1-a)\hat{\mu}_P^h}{\sqrt{\tilde{\sigma}_{E,h}^2/n_E + a^2\tilde{\sigma}_{R,h}^2/n_R + (1-a)^2\tilde{\sigma}_{P,h}^2/n_P}}$$

for $h = SHT, SRI, SAI$, which are asymptotically distributed as the standard normal distribution under $\tilde{H}_0$ as $\min\{n_E, n_R, n_P\} \to \infty$.

Note that the asymptotic properties of the parameter estimators and test statistics presented above only hold as $n_\ell \to \infty$ ($\ell = E, R, P$). However, for the finite samples, before using asymptotic normality of the estimators $\hat{\boldsymbol{\mu}}_\ell^h$ ($\ell = E, R, P$) and test statistics T_W^h ($h = SHT, SRI, SAI$), one should utilize the concept of goodness-of-fit tests [46,47] to check the plausibility of their normality assumption.

5. Simulation Study

In this section, simulation studies were conducted to assess the finite sample performance of the proposed test procedures in terms of empirical type I error rates and empirical powers under four missingness data mechanisms.

For $\ell = E, R, P$, the data $\{\boldsymbol{X}_{\ell i} : i = 1, \cdots, n_\ell\}$ were independently generated from the multivariate normal distribution, i.e., $\boldsymbol{X}_{\ell i} = (Z_{\ell i}, U_{\ell i}) \overset{\text{i.i.d}}{\sim} \mathcal{N}(\boldsymbol{\xi}, \boldsymbol{\Sigma})$, and the data $\{Y_{\ell i} : i = 1, \cdots, n_\ell\}$ were independently generated by $Y_{\ell i} = \boldsymbol{X}_{\ell i}^\top \boldsymbol{\beta}_\ell + \varepsilon_{\ell i}$, where $\varepsilon_{\ell i}$'s were independently sampled from the following normal distributions (denoted as 'scenario (A)'): $\varepsilon_{Ei} \sim \mathcal{N}(a\mu_R + (1-a)\mu_P, 0.34)$, $\varepsilon_{Ri} \sim \mathcal{N}(\mu_R, 0.37)$ and $\varepsilon_{Pi} \sim \mathcal{N}(\mu_P, 0.2)$ with $a = 0.8$, which was the three-arm "gold threshold" recommended in the considered literature [18]. The true values of $\boldsymbol{\xi}$, $\boldsymbol{\Sigma}$, $\boldsymbol{\beta}_E$, $\boldsymbol{\beta}_R$, $\boldsymbol{\beta}_P$, μ_R and μ_P were taken as $\boldsymbol{\xi} = (0.0, 0.0)^\top$, $\boldsymbol{\Sigma} = \text{diag}(0.25, 0.25)$, $\boldsymbol{\beta}_E = (1.0, 1.0)^\top$, $\boldsymbol{\beta}_R = (1.0, 1.1)^\top$, $\boldsymbol{\beta}_P = (0.5, 0.5)^\top$, $\mu_R = 1.1$ and $\mu_P = 0.6$, respectively, which were only chosen as an illustration of the proposed methodologies. For comparison with the cases used widely or always justified, we considered the following two scenarios: (B) $\varepsilon_{Ei} \sim \mathcal{N}(a\mu_R + (1-a)\mu_P, 0.8)$, $\varepsilon_{Ri} \sim \mathcal{N}(\mu_R, 0.8)$ and $\varepsilon_{Pi} \sim \mathcal{N}(\mu_P, 0.5)$; (C) $\varepsilon_{Ei} \sim \mathcal{N}(a\mu_R + (1-a)\mu_P, 1.0)$, $\varepsilon_{Ri} \sim \mathcal{N}(\mu_R, 1.0)$ and $\varepsilon_{Pi} \sim \mathcal{N}(\mu_P, 0.5)$. Under the above specified setting, we have $\psi(\boldsymbol{\mu}) = 0$, where $\boldsymbol{\mu} = \{\mu_E, \mu_R, \mu_P\}$. That is, the data $\{(\boldsymbol{X}_{\ell i}, Y_{\ell i}) : \ell = E, R, P, i = 1, \cdots, n_\ell\}$ were independently generated from the null hypothesis $\tilde{H}_0$, and were used to compute empirical type I error rates. To compute empirical powers, the data $\{(\boldsymbol{X}_{\ell i}, Y_{\ell i}) : \ell = E, R, P, i = 1, \cdots, n_\ell\}$ were independently generated with the above presented settings, except for $a = (\mu_E - \mu_P)/(\mu_R - \mu_P) > 0.8$, which implied that the data $\{(\boldsymbol{X}_{\ell i}, Y_{\ell i}) : \ell = E, R, P, i = 1, \cdots, n_\ell\}$ were sampled from the alternative hypothesis $\tilde{H}_1 : \psi(\mu) > 0$.

To create missing data for $Y_{\ell i}$, we assumed that the missing indicators $D_{\ell i}$'s were independently generated from the Bernoulli distribution with the respondent probability $\pi_{\ell i}$ for $\ell = E, R, P$. Here, we considered the following respondent probabilities for the reference treatment and placebo:

$$\text{logit}(\pi_{Ri}) = 1.6 + 0.2Z_{Ri} - 0.15Y_{Ri}, \quad \text{logit}(\pi_{Pi}) = 1.5 + 0.15Z_{Pi} - 0.18Y_{Pi},$$

which indicated that missingness data mechanisms were non-ignorable, and the following four respondent probabilities for experimental treatment:

Case E1: $\text{logit}(\pi_{Ei}) = \alpha_{E0} + \alpha_{E1} Z_{Ei}$, which led to a MAR missingness data mechanism, where the true values of α_{E0} and α_{E1} were taken to be 1.3 and 0.1, respectively.

Case E2: $\text{logit}(\pi_{Ei}) = \alpha_{E0} + \alpha_{E1} Z_{Ei} + \gamma_E Y_{Ei}$, which resulted in a non-ignorable missingness data mechanism, where the true values of α_{E0}, α_{E1} and γ_E were taken as 1.3, 0.1 and -0.1, respectively.

Case E3: $\text{logit}(\pi_{Ei}) = \alpha_{E0} + \alpha_{E1} \sin(Z_{Ei}) + \gamma_E Y_{Ei}$, which yielded a non-linear non-ignorable missingness data mechanism with respect to Z_{Ei}, where the true values of α_{E0}, α_{E1} and γ_E were set to be 1.3, 0.1 and 0.12, respectively.

Case E4: $\text{logit}(\pi_{Ei}) = \alpha_{E0} + \alpha_{E1} Z_{Ei} + \gamma_E Z_{Ei} Y_{Ei}$, which implied a non-linear non-ignorable missingness data mechanism with an interaction of Z_{Ei} and Y_{Ei}, where the true values of α_{E0}, α_{E1} and γ_E were set as 1.3, 0.1 and -0.1, respectively.

The titling parameters γ_ℓ corresponding to $Y_{\ell i}$ ($\ell = E, R, P$) were set to be roughly -0.2 for showing a moderately negative effect on the probability of the data observed, and $\alpha_{\ell 0}$ and $\alpha_{\ell 1}$ were chosen so that the average missing rates were roughly 25%. Case E1 was MAR, which was a special case of the considered missingness data mechanism model (4) with $\gamma_E = 0$ and was used to show that the proposed method can still capture missingness data characteristics even if the true missingness data mechanism was MAR; the other three missingness data mechanisms were non-ignorable and Case E2 satisfied the assumption of model (4), but Cases E3 and E4, which did not satisfy the assumed non-ignorable missingness data mechanism model (4), were used to show that the proposed test procedure was not sensitive to the assumed missingness data mechanism model (4). Here, we consider three balanced designs, i.e., $n_E = n_R = n_P = n$ with $n = 50, 100, 150$ for three scenarios, and the following unbalanced designs with the allocation ratios taken as 2:2:1, 3:3:1, 4:4:1, 2:1:1, 3:2:1, 4:3:1, 4:2:1, 3:1:1, 4:1:1 for Scenario (A) and 2:1:1, 2:2:1, 3:2:1 for Scenarios (B) and (C). The total sample sizes $N = n_E + n_R + n_P$ were set as 200 and 500 for Scenario (A) and 200, 300 and 400 for Scenarios (B) and (C), with a significance level $\alpha = 5\%$ for the three scenarios.

The average missing rates for the experimental, reference and placebo treatments among the 1000 replications were roughly 23.15%, 19.23% and 21.08%, respectively.

For each of the settings described above, we generated 1000 Monte Carlo samples. To evaluate the accuracy of the mean function estimates $\hat{m}^0_{\ell i}(\gamma_\ell; z_{\ell i})$ and the propensity score function estimates $\hat{\pi}_\ell(\boldsymbol{\eta}_\ell; z_{\ell i}, y_{\ell i})$, we took the Gaussian kernel function with $K(Z_\ell) = (2\pi)^{-1/2} \exp(-Z_\ell^2/2)$ and set the bandwidths h_ℓ as $\hat{\sigma}_{Z_\ell} n_\ell^{-1/3}$, where $\hat{\sigma}_{Z_\ell}$ was the standard deviation of observations $\{Z_{\ell i} : i = 1, \ldots, n_\ell\}$ for $\ell = E, R, P$. To compute the estimated asymptotic variances of $\hat{\mu}^k_\ell$, we conducted 100 bootstrap replications.

Empirical type I error rates for 1000 replications in Scenario (A) are given in Table 1 for balanced designs with the above considered four missingness data mechanisms and Table 2 for unbalanced designs with only the missingness data mechanism E2. To save space, we moved the corresponding results in Scenarios (B) and (C) to Tables A1 and A2 in the Appendix A. Examination of Tables 1, 2, A1 and A2 showed that (i) the proposed three statistics for testing $\tilde{H}_0$ have similar performance because their type I error rates are quite close to the pre-given significance level for all the considered cases, which is consistent with the theoretical properties presented in Theorems 1 and 2; (ii) the proposed three statistics for testing $\tilde{H}_0$ performed better than the CC method regardless of the sample sizes, missingness data mechanisms, balanced and unbalanced designs, and the variances of the treatment effects in that the type I error rates of the former were closer to the pre-given significance level than those for the latter; (iii) the type I error rate increased as the sample size increased for the CC method, which was consistent with the observations of Cook and Zea [48]; (iv) empirical type I error rates were not sensitive to the balanced or unbalanced designs.

Table 1. Empirical type I error rates for balanced designs in the first simulation study.

	n = 50				n = 100				n = 150			
Case	**SHT**	**SRI**	**SAI**	**CC**	**SHT**	**SRI**	**SAI**	**CC**	**SHT**	**SRI**	**SAI**	**CC**
E1	0.048	0.051	0.053	0.057	0.048	0.050	0.052	**0.074**	0.050	0.051	0.050	**0.083**
E2	0.053	0.055	0.055	0.052	0.048	0.049	0.050	0.055	0.054	0.055	0.055	**0.092**
E3	0.049	0.055	0.054	**0.070**	0.053	0.053	0.058	**0.095**	0.052	0.054	0.054	**0.117**
E4	0.051	0.053	0.052	**0.061**	0.048	0.050	0.053	**0.086**	0.047	0.051	0.049	**0.088**

Note: SHT, SRI, SAI and CC denote Wald-type test approaches based on IPW, regression imputation, AIPW and CC, respectively.

Table 2. Empirical type I error rates for unbalanced designs in the first simulation study.

	N = 200			N = 500		
n_E:n_R:n_P	**SHT**	**SRI**	**SAI**	**SHT**	**SRI**	**SAI**
2:2:1	0.046	0.048	0.048	0.049	0.050	0.051
3:3:1	0.053	0.053	0.051	0.045	0.046	0.046
4:4:1	0.053	0.054	0.058	0.045	0.047	0.050
2:1:1	0.048	0.051	0.050	0.045	0.044	0.045
3:2:1	0.055	0.059	0.060	0.047	0.049	0.046
4:3:1	0.038	0.044	0.047	0.054	0.055	0.055
4:2:1	0.048	0.051	0.053	0.063	0.064	0.059
3:1:1	0.055	0.056	0.057	0.052	0.055	0.056
4:1:1	0.050	0.057	0.058	0.051	0.052	0.054

We computed empirical powers against $a = (\mu_E - \mu_P)/(\mu_R - \mu_P)$, the sample sizes, the treatment effects, and the alpha and gamma parameters for missingness data mechanism models E1–E4 when the null hypothesis was not true. To save space, we only present empirical powers against the sample size in Figure 1 for balanced design and Figure 2 for unbalanced design (i.e., n_E:n_R:n_P = 2:1:1) under the considered four missingness data mechanism models. Other results are given in Figures A1–A8 in the Appendix A. Inspection of these figures showed that (i) empirical power increases as a or the sample size n increases, regardless of the missingness data mechanisms and balanced/unbalanced designs and the considered four tests; (ii) empirical power slightly increased as α_{E1} increased regardless of the missingness data mechanisms and balanced/unbalanced designs and the considered SHT, SRI and SAI tests, while the empirical power for the CC method showed an increasing tendency as α_{E1} increased for missingness data mechanisms E3 and E4, which might be explained by non-linear non-ignorable missing data; (iii) the empirical powers with the proposed three test statistics were larger than those with the CC method for non-ignorable missing data (i.e., E2–E4), regardless of the sample sizes, a, treatment effects, α_{E1} and γ; (iv) the observation that the CC method had a slightly larger empirical power than the three tests considered might be explained by its inflated type I error.

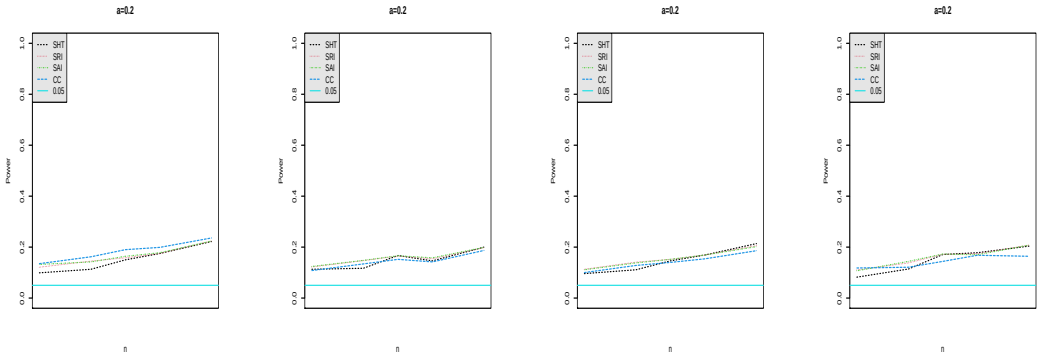

Figure 1. SHT, SRI, SAI and CC represent empirical powers evaluated from IPW, regression imputation, AIPW and CC methods against the sample size n under balanced design with missingness data mechanism models E1 (**left panel**), E2 (**left second panel**), E3 (**right second panel**) and E4 (**right panel**) for $a = 0.2$.

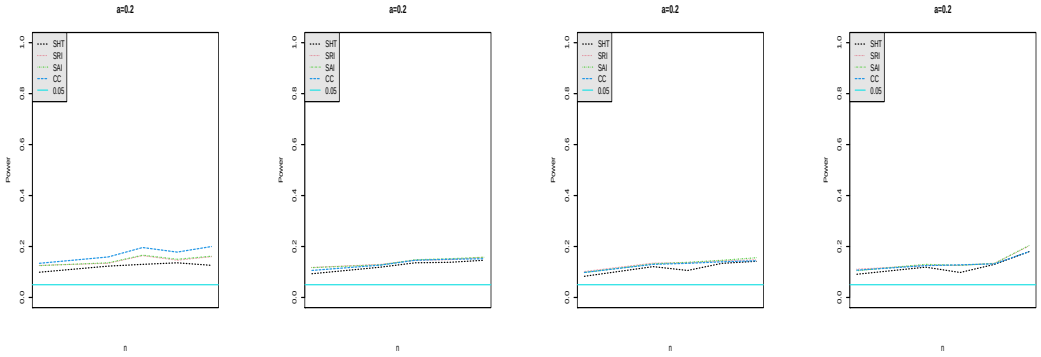

Figure 2. SHT, SRI, SAI and CC represent empirical powers evaluated from IPW, regression imputation, AIPW and CC methods against the sample size n under unbalanced design (i.e., n_E:n_R:n_P = 2:1:1), with missingness data mechanism models E1 (**left panel**), E2 (**left second panel**), E3 (**right second panel**) and E4 (**right panel**) for $a = 0.2$.

To investigate the effect of the amount of departure from the MAR mechanism (i.e., the change in γ_E) on type I error rates under the missingness data mechanism model E2, with the same values of α_{E0} and α_{E1} (i.e., $\alpha_{E0} = 1.3$ and $\alpha_{E1} = 0.1$) as those given in the first simulation study, we conducted a second simulation study. In this simulation study, 1000 Monte Carlo datasets $\{(\boldsymbol{X}_{\ell i}, Y_{\ell i}, D_{\ell i}) : i = 1, \cdots, n_\ell\}$ were independently generated, as in the first simulation study, with $\gamma_E = -0.2, -0.1, -0.05, 0, 0.05, 0.1$. Empirical type I error rates for the balanced design with the sample sizes $n = 50$ and 150 are given in Table 3. Inspection of Table 3 yielded that (i) statistics with the IPW, regression imputation and AIPW methods behaved better than those with the CC method because the type I error rates for the former were closer to the pre-given significance level than those for the latter, regardless of the values of γ_E and the sample sizes; and (ii) statistics with the IPW, regression imputation and AIPW methods were not sensitive to γ_E.

Table 3. Sensitivity analysis of the proposed test statistics in the second simulation study.

	$n = 50$				$n = 150$			
γ_E	**SHT**	**SRI**	**SAI**	**CC**	**SHT**	**SRI**	**SAI**	**CC**
−0.2	0.038	0.046	0.044	0.033	0.051	0.050	0.051	0.033
−0.1	0.053	0.055	0.055	0.052	0.054	0.055	0.055	**0.092**
−0.05	0.041	0.045	0.046	**0.066**	0.045	0.048	0.047	**0.061**
0.0	0.048	0.051	0.053	0.057	0.050	0.051	0.050	**0.083**
0.05	0.050	0.051	0.052	**0.065**	0.054	0.055	0.055	**0.092**
0.1	0.047	0.051	0.056	**0.087**	0.053	0.056	0.055	**0.111**

6. An Example

In this section, a real example described in the Introduction is used to illustrate the proposed methodologies. In this dataset, we regarded zidovudine plus 400 mg of didanosine plus 400 mg of nevirapine as the experimental treatment with 330 patients, zidovudine plus 2.25 mg of zalcitabine as the reference treatment with 324 patients, and zidovudine alternating monthly with 400 mg didanosine as the placebo with 325 patients, respectively. CD4 counts were scheduled to be collected at baseline and eight-week intervals during the follow-up. Due to mistimed measurements, CD4 count data were subject to missingness, which led to unbalanced designs. There were 94 patient dropouts at the interval (4,12] among the 330 patients, 97 patient dropouts at the interval (4,12] and two patient dropouts at baseline among the 324 patients, and 95 patient dropouts at the interval (4,12] and five patient dropouts at baseline among the 325 patients. As an illustration, we took log(CD4 count at baseline + 1) as the dropout instrument variable Z, and only considered the data at the interval (4,12], i.e., treatment endpoints were CD4 counts at the interval (4,12], which led to $n_E = 328$, $n_R = 313$ and $n_P = 316$, whose average missing rates were 29.74%, 30.99%, and 28.65%, respectively. The dataset was obtained from the R package "ALA". Our main purpose was to test the NI of the experimental treatment to the reference treatment in the considered three-arm design. To this end, we took the fraction margin as $\delta = g(\mu_R - \mu_P)$ with $g = 0.2$, which led to $a = 0.8$, i.e., the experimental treatment achieved more than 80 percent of the reference treatment compared with the placebo to be claimed as NI.

To compute $\hat{m}^0_{\ell i}(\hat{\gamma}_\ell; Z_{\ell i})$ and $\hat{\pi}_\ell(\hat{\boldsymbol{\eta}}_\ell; Z_{\ell i}, Y_{\ell i})$, we took the Gaussian kernel function as $K(x_\ell) = (2\pi)^{-1/2}\exp(-x_\ell^2/2)$ and set the bandwidth h_ℓ to be $\hat{\sigma}_{X_\ell} n_\ell^{-1/3}$, where $\hat{\sigma}_{X_\ell}$ was the standard deviation of $X_{\ell i}$'s. The p-values for testing $\tilde{H}_0$ were 0.0037, 0.0035, 0.0033 and 0.0136 for the Wald-type statistics with the IPW, regression imputation, AIPW and CC methods, respectively, which indicated the NI of the experimental treatment to the reference treatment was at the 5% significance level.

7. Conclusions

This paper considers the non-inferiority assessment problem of an experimental treatment to a reference treatment in a three-arm trial with non-ignorable missing data. A logistic regression was employed to specify the non-ignorable missing endpoint mechanism. Three methods, including the IPW, imputation regression and AIPW methods, were proposed to estimate the treatment efficacy for the known and unknown propensity score functions. The asymptotic properties of estimators for treatment efficacy were established under some regularity conditions. Based on these asymptotic properties, three Wald-type statistics for testing the NI of the experimental treatment to the reference treatment were constructed. Simulation studies indicated that the proposed test procedures behaved better than those with complete-case data in terms of type I error rates and powers, i.e., the type I error rates for the former were closer to the pre-given significance level than those for the latter and the powers for the former were larger than those for the latter; the proposed test procedures were not sensitive to misspecified missingness data mechanisms.

Author Contributions: W.L. carried out the simulation studies, the statistical analysis and drafted the manuscript; Y.Z. drafted the manuscript and conducted the statistical analysis; N.T. conceived the research questions and idea, developed the methods and revised the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: The authors are very grateful to the Editor, the Associate Editor and the three anonymous referees for their valuable comments that significantly improved this work. This work was supported by grants from the National Natural Science Foundation of China (No. 12271472), the National Key R&D Program of China (No. 2022YFA1003701), and the China Postdoctoral Science Foundation (No. 2021M702778).

Data Availability Statement: The datasets generated and analysed during the current study are available from the corresponding author on reasonable request.

Conflicts of Interest: The authors declare no competing interest.

Appendix A

Appendix A.1. Regularity Conditions

To obtain asymptotic properties of $\hat{\mu}_\ell$ for $\ell = E, R, P$, we need the following regularity conditions.

Assumption A1. *The true respondent model given in Equation (4) satisfies (i) that there exists a true value $\boldsymbol{\eta}_\ell^0$ of $\boldsymbol{\eta}_\ell$ such that $E\{s(\boldsymbol{\eta}_\ell^0; \mathbf{Z}_\ell, Y_\ell)\} = 0$; (ii) for $\boldsymbol{\eta}_\ell$, in a neighborhood of $\boldsymbol{\eta}_\ell^0$, $E\{||s(\boldsymbol{\eta}_\ell^0; \mathbf{Z}_\ell, Y_\ell)||^2\} < \infty$ and $E\{\partial s(\boldsymbol{\eta}_\ell^0; \mathbf{Z}_\ell, Y_\ell)/\partial \eta_\ell^T\}$ exists and is nonsingular.*

Assumption A2. *(i) The marginal probability density function $f(z)$ of the random variable z is bounded away from ∞ in the support of z, and the second derivative of $f(z)$ in z is continuous and bounded; (ii) The respondent probabilities $\pi_{\ell i}(\boldsymbol{\eta}_\ell; \mathbf{Z}_{\ell i}, Y_{\ell i})$ satisfy $\min_{1\le i\le n_\ell} \pi_{\ell i}(\boldsymbol{\eta}_\ell; \mathbf{Z}_{\ell i}, Y_{\ell i}) \ge c_0$ a.s. for some positive constant c_0.*

Assumption A3. *The kernel function $K_\ell(\cdot)$ satisfies (i) it is bounded and has compact support; (ii) it is symmetric with $\int \omega^2 K_\ell(\omega) d\omega < \infty$; (iii) $K_\ell(\cdot) \ge D_\ell$ for $D_\ell > 0$ in some closed interval centered at zero; (iv) $n_\ell h_\ell \to \infty$ and $n_\ell h_\ell^4 \to 0$ as $n_\ell \to \infty$.*

Remark A1. *Assumption A1 is used to establish asymptotic normality of $\hat{\boldsymbol{\eta}}_\ell$. Assumption A2 is commonly adopted in the missing data literature. Assumption A3 is a standard assumption for the kernel regression method.*

Proof of Theorem 1. By the definition of $\hat{\mu}_\ell^{HT}$, we have the following decomposition

$$\begin{aligned}\sqrt{n_\ell}(\hat{\mu}_\ell^{HT} - \mu_\ell^0) =& n_\ell^{-1/2} \sum_{i=1}^{n_\ell} \left\{ \frac{D_{\ell i} Y_{\ell i}}{\pi_{\ell i}(\boldsymbol{\eta}_\ell; \mathbf{Z}_{\ell i}, Y_{\ell i})} - \mu_\ell^0 \right\} \\ =& n_\ell^{-1/2} \sum_{i=1}^{n_\ell} \frac{D_{\ell i}\{Y_{\ell i} - m_{\ell i}^0(\gamma_\ell; \mathbf{Z}_{\ell i})\}}{\pi_{\ell i}(\boldsymbol{\eta}_\ell; \mathbf{Z}_{\ell i}, Y_{\ell i})} + n_\ell^{-1/2} \sum_{i=1}^{n_\ell} \left\{ m_{\ell i}^0(\gamma_\ell; \mathbf{Z}_{\ell i}) - \mu_\ell^0 \right\} \\ &+ n_\ell^{-1/2} \sum_{i=1}^{n_\ell} \left\{ \frac{D_{\ell i}}{\pi_{\ell i}(\boldsymbol{\eta}_\ell; \mathbf{Z}_{\ell i}, Y_{\ell i})} - 1 \right\} m_{\ell i}^0(\gamma_\ell; \mathbf{Z}_{\ell i}) \triangleq H_{1n_\ell} + H_{2n_\ell} + H_{3n_\ell}.\end{aligned}$$

By the law of large numbers, it is easily shown that $H_{3n_\ell} = o_p(1)$. Thus, combining the above results leads to

$$\sqrt{n_\ell}(\hat{\mu}_\ell^{HT} - \mu_\ell^0) = n_\ell^{-1/2} \sum_{i=1}^{n_\ell} \frac{D_{\ell i}\{Y_{\ell i} - m_{\ell i}^0(\gamma_\ell; \mathbf{Z}_{\ell i})\}}{\pi_{\ell i}(\boldsymbol{\eta}_\ell; \mathbf{Z}_{\ell i}, Y_{\ell i})} + n_\ell^{-1/2} \sum_{i=1}^{n_\ell} \left\{ m_{\ell i}^0(\gamma_\ell; \mathbf{Z}_{\ell i}) - \mu_\ell^0 \right\} + o_p(1).$$

Let $\tau_{\ell i} = m^0_{\ell i}(\gamma_\ell; \mathbf{Z}_{\ell i}) + D_{\ell i}\pi^{-1}_{\ell i}(\boldsymbol{\eta}^0_\ell)\{y_{\ell i} - m^0_{\ell i}(\gamma_\ell; \mathbf{Z}_{\ell i})\}$. Then, we have

$$\sqrt{n_\ell}\left(\hat{\mu}^{HT}_\ell - \mu^0_\ell\right) \xrightarrow{\mathcal{L}} N\left(0, \sigma^2_\ell\right) \quad as \quad n_\ell \to \infty\,, \text{ for } \ell = E, R, P,$$

where $\sigma^2_\ell = \text{var}(\tau_{\ell i})$. It is easily shown that $E\left\{D_{\ell i}\pi_{\ell i}(\boldsymbol{\eta}_\ell; \mathbf{Z}_{\ell i}, Y_{\ell i})^{-1}\{Y_{\ell i} - m^0_{\ell i}(\gamma_\ell; \mathbf{Z}_{\ell i})\}\right\} = 0$, and $E\{m^0_{\ell i}(\gamma_\ell; \mathbf{Z}_{\ell i}) - \mu^0_\ell\} = 0$. Since $m^0_{\ell i}(\gamma_\ell; \mathbf{Z}_{\ell i})$ is independent of $Y_{\ell i} - m^0_{\ell i}(\gamma_\ell; \mathbf{Z}_{\ell i})$, we have $\sigma^2_\ell = E\{\pi_{\ell i}(Z_{\ell i}, Y_{\ell i}; \boldsymbol{\eta}^0_\ell)^{-1}[Y_{\ell i} - m^0_{\ell i}(\gamma_\ell; \mathbf{Z}_{\ell i}]^2\} + E\left[m^0_{\ell i}(\gamma_\ell; \mathbf{Z}_{\ell i})\right]^2$.

Next, we show the asymptotic property of $\hat{\mu}^{RI}_\ell$. By the definition of $\hat{\mu}^{RI}_\ell$, we obtain

$$\begin{aligned}
\sqrt{n_\ell}(\hat{\mu}^{RI}_\ell - \mu^0_\ell) &= n^{-1/2}_\ell \sum_{i=1}^{n_\ell}\left\{D_{\ell i}Y_{\ell i} + (1 - D_{\ell i})\hat{m}^0_{\ell i}(\gamma_\ell; \mathbf{Z}_{\ell i}) - \mu^0_\ell\right\} \\
&= n^{-1/2}_\ell \sum_{i=1}^{n_\ell} D_{\ell i}\left\{Y_{\ell i} - m^0_{\ell i}(\gamma_\ell; \mathbf{Z}_{\ell i})\right\} + n^{-1/2}_\ell \sum_{i=1}^{n_\ell}(1 - D_{\ell i})\left\{\hat{m}^0_{\ell i}(\gamma_\ell; \mathbf{Z}_{\ell i}) - m^0_{\ell i}(\gamma_\ell; \mathbf{Z}_{\ell i})\right\} \\
&\quad + n^{-1/2}_\ell \sum_{i=1}^{n_\ell}\left\{m^0_{\ell i}(\gamma_\ell; \mathbf{Z}_{\ell i}) - \mu^0_\ell\right\} \triangleq R_{1n_\ell} + R_{2n_\ell} + H_{2n_\ell}.
\end{aligned}$$

Using the similar arguments as given in Tang et al. [38], it is easily shown that

$$R_{2n_\ell} = n^{-1/2}_\ell \sum_{i=1}^{n_\ell} D_{\ell i}[1 - \pi_{\ell i}(\boldsymbol{\eta}^0_\ell; \mathbf{Z}_{\ell i}, Y_{\ell i})]\{Y_{\ell i} - m^0_{\ell i}(\gamma_\ell; \mathbf{Z}_{\ell i})\}/\pi_{\ell i}(\boldsymbol{\eta}^0_\ell; \mathbf{Z}_{\ell i}, Y_{\ell i}) + o_p(1).$$

Thus, we have

$$\sqrt{n_\ell}(\hat{\mu}^{RI}_\ell - \mu^0_\ell) = H_{1n_\ell} + H_{2n_\ell} + o_p(1) = \sqrt{n_\ell}(\hat{\mu}^{HT}_\ell - \mu^0_\ell) + o_p(1).$$

By the Slutsky Theorem and the asymptotic property of $\hat{\mu}^{HT}_\ell$, we obtain

$$\sqrt{n_\ell}\left(\hat{\mu}^{RI}_\ell - \mu^0_\ell\right) \xrightarrow{\mathcal{L}} N\left(0, \sigma^2_\ell\right) \quad as \quad n_\ell \to \infty\,, \text{ for } \ell = E, R, P,$$

where σ^2_ℓ is defined in the proof of the asymptotic properties of $\hat{\mu}^{HT}_\ell$.

Now, we prove the asymptotic properties of the estimator $\hat{\mu}^{AI}_\ell$. By the definition of the $\hat{\mu}^{AI}_\ell$, we obtain

$$\begin{aligned}
\sqrt{n_\ell}(\hat{\mu}^{AI}_\ell - \mu^0_\ell) &= n^{-1/2}_\ell \sum_{i=1}^{n_\ell}\left\{\frac{D_{\ell i}Y_{\ell i}}{\pi_{\ell i}(\boldsymbol{\eta}_\ell; \mathbf{Z}_{\ell i}, Y_{\ell i})} + \left(1 - \frac{D_{\ell i}}{\pi_{\ell i}(\boldsymbol{\eta}_\ell; \mathbf{Z}_{\ell i}, Y_{\ell i})}\right)\hat{m}^0_{\ell i}(\gamma_\ell; \mathbf{Z}_{\ell i}) - \mu^0_\ell\right\} \\
&= n^{-1/2}_\ell \sum_{i=1}^{n_\ell} \frac{D_{\ell i}}{\pi_{\ell i}(\boldsymbol{\eta}_\ell; \mathbf{Z}_{\ell i}, Y_{\ell i})}\left\{Y_{\ell i} - m^0_{\ell i}(\gamma_\ell; \mathbf{Z}_{\ell i})\right\} + n^{-1/2}_\ell \sum_{i=1}^{n_\ell}\left\{m^0_{\ell i}(\gamma_\ell; \mathbf{Z}_{\ell i}) - \mu^0_\ell\right\} \\
&\quad + n^{-1/2}_\ell \sum_{i=1}^{n_\ell}\left(1 - \frac{D_{\ell i}}{\pi_{\ell i}(\boldsymbol{\eta}_\ell; \mathbf{Z}_{\ell i}, Y_{\ell i})}\right)\left\{\hat{m}^0_{\ell i}(\gamma_\ell; \mathbf{Z}_{\ell i}) - m^0_{\ell i}(\gamma_\ell; \mathbf{Z}_{\ell i})\right\} \triangleq H_{1n_\ell} + H_{2n_\ell} + A_{n_\ell}.
\end{aligned}$$

Following a similar argument as given in the proof of Theorem 4 in Zhao et al. [49], we have $A_{n_\ell} = o_p(1)$. Combining the above results yields

$$\sqrt{n_\ell}(\hat{\mu}^{AI}_\ell - \mu^0_\ell) = H_{1n_\ell} + H_{2n_\ell} + o_p(1) = \sqrt{n_\ell}(\hat{\mu}^{HT}_\ell - \mu^0_\ell) + o_p(1).$$

Using the same arguments as given in the proof of the asymptotic properties for $\hat{\mu}^{RI}_\ell$, we obtain

$$\sqrt{n_\ell}\left(\hat{\mu}^{AI}_\ell - \mu^0_\ell\right) \xrightarrow{\mathcal{L}} N\left(0, \sigma^2_\ell\right) \quad as \quad n_\ell \to \infty\,, \text{ for } \ell = E, R, P,$$

where σ^2_ℓ is defined in the proof of the asymptotic properties of $\hat{\mu}^{HT}_\ell$. □

Proof of Theorem 2. We first consider the asymptotic properties of $\hat{\mu}_\ell^{SHT}$ based on the following form:

$$\sqrt{n_\ell}(\hat{\mu}_\ell^{SHT} - \mu_\ell^0) = n_\ell^{-1/2} \sum_{i=1}^{n_\ell} \Big\{ \frac{D_{\ell i} Y_{\ell i}}{\pi_{\ell i}(\hat{\boldsymbol{\eta}}_\ell; \mathbf{Z}_{\ell i}, Y_{\ell i})} - \mu_\ell^0 \Big\}.$$

Taking the Taylor expansion of $\pi_{\ell i}(\hat{\boldsymbol{\eta}}_\ell; \mathbf{Z}_{\ell i}, Y_{\ell i})$ at $\boldsymbol{\eta}_\ell$ yields

$$\begin{aligned} n_\ell^{-1/2} \sum_{i=1}^{n_\ell} \frac{D_{\ell i} Y_{\ell i}}{\pi_{\ell i}(\hat{\boldsymbol{\eta}}_\ell; \mathbf{Z}_{\ell i}, Y_{\ell i})} &= n_\ell^{-1/2} \sum_{i=1}^{n_\ell} \frac{D_{\ell i} Y_{\ell i}}{\pi_{\ell i}(\boldsymbol{\eta}_\ell; \mathbf{Z}_{\ell i}, Y_{\ell i})} \\ &\quad + n_\ell^{1/2} (\hat{\boldsymbol{\eta}}_\ell - \boldsymbol{\eta}_\ell)^\top \frac{1}{n_\ell} \sum_{i=1}^{n_\ell} D_{\ell i} Y_{\ell i} \frac{\partial}{\partial \boldsymbol{\eta}_\ell} \pi_{\ell i}^{-1}(\boldsymbol{\eta}_\ell; \mathbf{Z}_{\ell i}, Y_{\ell i}) \big|_{\boldsymbol{\eta}_\ell = \tilde{\boldsymbol{\eta}}_\ell} + o_p(1) \end{aligned}$$

where $\tilde{\boldsymbol{\eta}}_\ell$ lies in the line segment between $\hat{\boldsymbol{\eta}}_\ell$ and $\boldsymbol{\eta}_\ell$, $\partial \pi_{\ell i}^{-1}(\boldsymbol{\eta}_\ell; \mathbf{Z}_{\ell i}, Y_{\ell i}) / \partial \boldsymbol{\eta}_\ell \big|_{\boldsymbol{\eta}_\ell = \tilde{\boldsymbol{\eta}}_\ell} = \{1 - \pi_{\ell i}^{-1}(\boldsymbol{\eta}_\ell; \mathbf{Z}_{\ell i}, Y_{\ell i})\}(1, \mathbf{Z}_{\ell i}^\top, Y_{\ell i})^\top$. Following the arguments of the proof of Theorem 1, we get

$$\begin{aligned} n_\ell^{-1/2} \sum_{i=1}^{n_\ell} \Big\{ \frac{D_{\ell i} Y_{\ell i}}{\pi_{\ell i}(\boldsymbol{\eta}_\ell; \mathbf{Z}_{\ell i}, Y_{\ell i})} - \mu_\ell^0 \Big\} &= n_\ell^{-1/2} \sum_{i=1}^{n_\ell} \frac{D_{\ell i} \{ Y_{\ell i} - m_{\ell i}^0(\gamma_\ell; \mathbf{Z}_{\ell i}) \}}{\pi_{\ell i}(\boldsymbol{\eta}_\ell; \mathbf{Z}_{\ell i}, Y_{\ell i})} \\ &\quad + n_\ell^{-1/2} \sum_{i=1}^{n_\ell} \{ m_{\ell i}^0(\gamma_\ell; \mathbf{Z}_{\ell i}) - \mu_\ell^0 \} + o_p(1). \end{aligned}$$

Combining the above results yields

$$\begin{aligned} \sqrt{n_\ell}(\hat{\mu}_\ell^{SHT} - \mu_\ell^0) =& n_\ell^{-1/2} \sum_{i=1}^{n_\ell} \frac{D_{\ell i} \{ Y_{\ell i} - m_{\ell i}^0(\gamma_\ell; \mathbf{Z}_{\ell i}) \}}{\pi_{\ell i}(\boldsymbol{\eta}_\ell; \mathbf{Z}_{\ell i}, Y_{\ell i})} + n_\ell^{-1/2} \sum_{i=1}^{n_\ell} \{ m_{\ell i}^0(\gamma_\ell; \mathbf{Z}_{\ell i}) - \mu_\ell^0 \} \\ &+ n_\ell^{1/2} (\hat{\boldsymbol{\eta}}_\ell - \boldsymbol{\eta}_\ell)^\top \frac{1}{n_\ell} \sum_{i=1}^{n_\ell} D_{\ell i} Y_{\ell i} \Big(1 - \frac{1}{\pi_{\ell i}(\tilde{\boldsymbol{\eta}}_\ell; \mathbf{Z}_{\ell i}, Y_{\ell i})}\Big)(1, \mathbf{Z}_{\ell i}^\top, Y_{\ell i})^\top + o_p(1) \\ =& n_\ell^{-1/2} \sum_{i=1}^{n_\ell} \left\{ \frac{D_{\ell i} \{ Y_{\ell i} - m_{\ell i}^0(\gamma_\ell; \mathbf{Z}_{\ell i}) \}}{\pi_{\ell i}(\boldsymbol{\eta}_\ell; \mathbf{Z}_{\ell i}, Y_{\ell i})} + \Big\{ m_{\ell i}^0(\gamma_\ell; \mathbf{Z}_{\ell i}) - \mu_\ell^0 \Big\} \right\} \\ &+ n_\ell^{1/2} (\hat{\boldsymbol{\eta}}_\ell - \boldsymbol{\eta}_\ell)^\top H_{\ell, SHT} + o_p(1) \\ =& n_\ell^{-1/2} \sum_{i=1}^{n_\ell} e_{\ell, SHTi} + o_p(1), \end{aligned}$$

where

$$e_{\ell, SHTi} = \{ D_{\ell i} \pi_{\ell i}^{-1}(\boldsymbol{\eta}_\ell; \mathbf{Z}_{\ell i}, Y_{\ell i}) \{ Y_{\ell i} - m_{\ell i}^0(\gamma_\ell; \mathbf{Z}_{\ell i}) \} + m_{\ell i}^0(\gamma_\ell; \mathbf{Z}_{\ell i}) - \mu_\ell^0 + \mathcal{I}_{22\ell}^{-1} s_{\ell i}(\boldsymbol{\eta}_\ell) H_{\ell, SHT} \},$$

$s_{\ell i}(\boldsymbol{\eta}_\ell)$ is the ith term in Equation (11), $H_{\ell, SHT} = E[\{ \pi_\ell(\boldsymbol{\eta}_\ell; \mathbf{Z}_\ell, Y_\ell) - 1 \} Y_\ell (1, \mathbf{Z}_\ell^\top, Y_\ell)^\top]$. By the Slutsky Theorem and the asymptotic property of $\hat{\mu}_\ell^{HT}$, it is easily shown that

$$\sqrt{n_\ell} \Big(\hat{\mu}_\ell^{SHT} - \mu_\ell^0 \Big) \xrightarrow{\mathcal{L}} N\Big(0, \sigma_{\ell, SHT}^2\Big) \quad as \quad n_\ell \to \infty, \text{ for } \ell = E, R, P,$$

where $\sigma_{\ell, SHT}^2 = \mathrm{Var}(e_{\ell, SHTi})$.

Now, we show the asymptotic property of $\hat{\mu}_\ell^{SRI}$ for unknown $\boldsymbol{\eta}_\ell$. By the definition of $\hat{\mu}_\ell^{SRI}$, we have the following form:

$$\sqrt{n_\ell}(\hat{\mu}_\ell^{SRI} - \mu_\ell^0) = n_\ell^{-1/2} \sum_{i=1}^{n_\ell} \Big\{ D_{\ell i} Y_{\ell i} + (1 - D_{\ell i}) \hat{m}_{\ell i}^0(\hat{\gamma}_\ell; \mathbf{Z}_{\ell i}) - \mu_\ell^0 \Big\}.$$

Taking the Taylor expansion of $\hat{m}_{\ell i}^0(\hat{\gamma}_\ell; \mathbf{Z}_{\ell i})$ at $\boldsymbol{\eta}_\ell$ yields

$$\begin{aligned}\sum_{i=1}^{n_\ell}(1-D_{\ell i})\hat{m}^0_{\ell i}(\hat{\gamma}_\ell;\mathbf{Z}_{\ell i}) &= \sum_{i=1}^{n_\ell}(1-D_{\ell i})\hat{m}^0_{\ell i}(\gamma_\ell;\mathbf{Z}_{\ell i})\\ &+ n_\ell^{1/2}(\hat{\boldsymbol{\eta}}_\ell-\boldsymbol{\eta}_\ell)^\top\frac{1}{n_\ell}\sum_{i=1}^{n_\ell}(1-D_{\ell i})\frac{\partial}{\partial\boldsymbol{\eta}_\ell}m^0_{\ell i}(\gamma_\ell;\mathbf{Z}_{\ell i})\Big|_{\boldsymbol{\eta}_\ell=\tilde{\boldsymbol{\eta}}_\ell}+o_p(1).\end{aligned}$$

Using the conclusion given in the proof of Theorem 1, we can easily get $\sqrt{n_\ell}(\hat{\mu}_\ell^{RI}-\mu_\ell^0)=\sqrt{n_\ell}(\hat{\mu}_\ell^{HT}-\mu_\ell^0)+o_p(1)$. Combining the above results yields

$$\begin{aligned}\sqrt{n_\ell}(\hat{\mu}_\ell^{SRI}-\mu_\ell^0) =& n_\ell^{-1/2}\sum_{i=1}^{n_\ell}\Big\{D_{\ell i}Y_{\ell i}+(1-D_{\ell i})\hat{m}^0_{\ell i}(\gamma_\ell;\mathbf{Z}_{\ell i})-\mu_\ell^0\Big\}\\ &+n_\ell^{1/2}(\hat{\boldsymbol{\eta}}_\ell-\boldsymbol{\eta}_\ell)^\top\frac{1}{n_\ell}\sum_{i=1}^{n_\ell}(1-D_{\ell i})Y_{\ell i}\frac{\partial}{\partial\boldsymbol{\eta}_\ell}\hat{m}^0_{\ell i}(\gamma_\ell;\mathbf{Z}_{\ell i})\Big|_{\boldsymbol{\eta}_\ell=\tilde{\boldsymbol{\eta}}_\ell}\\ =& n_\ell^{-1/2}\sum_{i=1}^{n_\ell}\frac{D_{\ell i}\{Y_{\ell i}-m^0_{\ell i}(\gamma_\ell;\mathbf{Z}_{\ell i})\}}{\pi_{\ell i}(\boldsymbol{\eta}_\ell;\mathbf{Z}_{\ell i},Y_{\ell i})}+n_\ell^{-1/2}\sum_{i=1}^{n_\ell}\Big\{m^0_{\ell i}(\gamma_\ell;\mathbf{Z}_{\ell i})-\mu_\ell^0\Big\}\\ &+n_\ell^{1/2}(\hat{\boldsymbol{\eta}}_\ell-\boldsymbol{\eta}_\ell)^\top\mathcal{I}_{22\ell}^{-1}s_{\ell i}(\boldsymbol{\eta}_\ell)H_{\ell,SRI}+o_p(1)\\ =& n_\ell^{-1/2}\sum_{i=1}^{n_\ell}e_{\ell,SRIi}+o_p(1),\end{aligned}$$

where $\tilde{\boldsymbol{\eta}}_\ell$ lies in the line segment between $\hat{\boldsymbol{\eta}}_\ell$ and $\boldsymbol{\eta}_\ell$, $e_{\ell,SRIi}=D_{\ell i}\pi_{\ell i}^{-1}(\boldsymbol{\eta}_\ell;\mathbf{Z}_{\ell i},Y_{\ell i})\{Y_{\ell i}-m^0_{\ell i}(\gamma_\ell;\mathbf{Z}_{\ell i})\}+m^0_{\ell i}(\gamma_\ell;\mathbf{Z}_{\ell i})-\mu_\ell^0+\mathcal{I}_{22\ell}^{-1}s_{\ell i}(\boldsymbol{\eta}_\ell)H_{\ell,SRI}$, $H_{\ell,SRI}=E\{(1-D_{\ell i})(0,\mathbf{0}_{p_\ell-1}^\top,(Y_{\ell i}-m^0_{\ell i}(\gamma_\ell;\mathbf{Z}_{\ell i}))^2)^\top\}$ and $\mathbf{0}_{p_\ell-1}$ is a $(p_\ell-1)\times 1$ zero vector and p_ℓ is the dimension of covariate $\mathbf{Z}_{\ell i}$. By the Slutsky Theorem and the asymptotic property of $\hat{\mu}_\ell^{SHT}$, it is easily shown that

$$\sqrt{n_\ell}\Big(\hat{\mu}_\ell^{SRI}-\mu_\ell^0\Big)\xrightarrow{\mathcal{L}}N\Big(0,\sigma^2_{\ell,SRI}\Big)\quad as\quad n_\ell\to\infty,\ \text{ for } \ell=E,R,P,$$

where $\sigma^2_{\ell,SRI}=\mathrm{Var}(e_{\ell,SRIi})$.

Now, we prove the asymptotic properties of $\hat{\mu}_\ell^{SAI}$ for unknown $\boldsymbol{\eta}_\ell$. Combining the above results and taking the Taylor expansion of $\hat{\mu}_\ell^{SAI}$ at $\boldsymbol{\eta}_\ell$, we obtain

$$\begin{aligned}\sqrt{n_\ell}(\hat{\mu}_\ell^{SAI}-\mu_\ell^0) =& n_\ell^{-1/2}\sum_{i=1}^{n_\ell}\Big\{\frac{D_{\ell i}Y_{\ell i}}{\pi_{\ell i}(\boldsymbol{\eta}_\ell;\mathbf{Z}_{\ell i},Y_{\ell i})}+\Big(1-\frac{D_{\ell i}}{\pi_{\ell i}(\boldsymbol{\eta}_\ell;\mathbf{Z}_{\ell i},Y_{\ell i})}\Big)\hat{m}^0_{\ell i}(\gamma_\ell;\mathbf{Z}_{\ell i})-\mu_\ell^0\Big\}\\ &+n_\ell^{1/2}(\hat{\boldsymbol{\eta}}_\ell-\boldsymbol{\eta}_\ell)^\top\frac{1}{n_\ell}\sum_{i=1}^{n_\ell}D_{\ell i}Y_{\ell i}\frac{\partial}{\partial\boldsymbol{\eta}_\ell}\pi_{\ell i}^{-1}(\boldsymbol{\eta}_\ell;\mathbf{Z}_{\ell i},Y_{\ell i})\big|_{\boldsymbol{\eta}_\ell=\tilde{\boldsymbol{\eta}}_\ell}\\ &+n_\ell^{1/2}(\hat{\boldsymbol{\eta}}_\ell-\boldsymbol{\eta}_\ell)^\top\frac{1}{n_\ell}\sum_{i=1}^{n_\ell}\hat{m}^0_{\ell i}(\gamma_\ell;\mathbf{Z}_{\ell i})\frac{\partial}{\partial\boldsymbol{\eta}_\ell}\Big(1-\frac{D_{\ell i}}{\pi_{\ell i}(\boldsymbol{\eta}_\ell;\mathbf{Z}_{\ell i},Y_{\ell i})}\Big)\Big|_{\boldsymbol{\eta}_\ell=\tilde{\boldsymbol{\eta}}_\ell}\\ &+n_\ell^{1/2}(\hat{\boldsymbol{\eta}}_\ell-\boldsymbol{\eta}_\ell)^\top\frac{1}{n_\ell}\sum_{i=1}^{n_\ell}\Big(1-\frac{D_{\ell i}}{\pi_{\ell i}(\boldsymbol{\eta}_\ell;\mathbf{Z}_{\ell i},Y_{\ell i})}\Big)\frac{\partial}{\partial\boldsymbol{\eta}_\ell}\hat{m}^0_{\ell i}(\gamma_\ell;\mathbf{Z}_{\ell i})\Big|_{\boldsymbol{\eta}_\ell=\tilde{\boldsymbol{\eta}}_\ell}\end{aligned}$$

where $\tilde{\boldsymbol{\eta}}_\ell$ lies in the line segment between $\hat{\boldsymbol{\eta}}_\ell$ and $\boldsymbol{\eta}_\ell$. Using the conclusion given in the proof of Theorem 1, we have

$$\sqrt{n_\ell}(\hat{\mu}_\ell^{AI}-\mu_\ell^0)=\sqrt{n_\ell}(\hat{\mu}_\ell^{HT}-\mu_\ell^0)+o_p(1),$$

$$\frac{1}{n_\ell}\sum_{i=1}^{n_\ell}\Big(1-\frac{D_{\ell i}}{\pi_{\ell i}(\boldsymbol{\eta}_\ell;\mathbf{Z}_{\ell i},Y_{\ell i})}\Big)\frac{\partial}{\partial\boldsymbol{\eta}_\ell}\hat{m}^0_{\ell i}(\gamma_\ell;\mathbf{Z}_{\ell i})|_{\boldsymbol{\eta}_\ell=\tilde{\boldsymbol{\eta}}_\ell}=o_p(1),$$

$$\frac{1}{n_\ell}\sum_{i=1}^{n_\ell}\hat{m}^0_{\ell i}(\gamma_\ell;\mathbf{Z}_{\ell i})\frac{\partial}{\partial\boldsymbol{\eta}_\ell}\Big(1-\frac{D_{\ell i}}{\pi_{\ell i}(\boldsymbol{\eta}_\ell;\mathbf{Z}_{\ell i},Y_{\ell i})}\Big)|_{\boldsymbol{\eta}_\ell=\tilde{\boldsymbol{\eta}}_\ell}=M_{\ell SAI}+o_p(1),$$

where $M_{\ell,SAI} = E\{(1-\pi_{\ell i}(\boldsymbol{\eta}_\ell;\boldsymbol{Z}_{\ell i},Y_{\ell i}))m^0_{\ell i}(\gamma_\ell;\boldsymbol{Z}_{\ell i})(1,\boldsymbol{Z}^\top_{\ell i},Y_{\ell i})^\top\}$. Combining the above results leads to

$$\begin{aligned}\sqrt{n_\ell}(\hat{\mu}^{SAI}_\ell-\mu^0_\ell) =& n_\ell^{-1/2}\sum_{i=1}^{n_\ell}\left[\frac{D_{\ell i}\{Y_{\ell i}-m^0_{\ell i}(\gamma_\ell;\boldsymbol{Z}_{\ell i})\}}{\pi_{\ell i}(\boldsymbol{\eta}_\ell;\boldsymbol{Z}_{\ell i},Y_{\ell i})}+\left\{m^0_{\ell i}(\gamma_\ell;\boldsymbol{Z}_{\ell i})-\mu^0_\ell\right\}\right]\\ &+n_\ell^{1/2}(\hat{\boldsymbol{\eta}}_\ell-\boldsymbol{\eta}_\ell)^\top\mathcal{I}^{-1}_{22\ell}s_{\ell i}(\boldsymbol{\eta}_\ell)(H_{\ell,SHT}+M_{\ell,SAI})+o_p(1)\\ =& n_\ell^{-1/2}\sum_{i=1}^{n_\ell}e_{\ell,SAIi}+o_p(1),\end{aligned}$$

where $e_{\ell,SAIi} = D_{\ell i}\pi^{-1}_{\ell i}(\boldsymbol{\eta}_\ell;\boldsymbol{Z}_{\ell i},Y_{\ell i})\{Y_{\ell i}-m^0_{\ell i}(\gamma_\ell;\boldsymbol{Z}_{\ell i})\}+m^0_{\ell i}(\gamma_\ell;\boldsymbol{Z}_{\ell i})-\mu^0_\ell+\mathcal{I}^{-1}_{22\ell}s_{\ell i}(\boldsymbol{\eta}_\ell)$ $H_{\ell,SRI}$, $H_{\ell,SAI} = H_{\ell,SHT}+M_{\ell,SAI}$. By the Slutsky Theorem and the asymptotic property of $\hat{\mu}^{SHT}_\ell$, it is easily shown that

$$\sqrt{n_\ell}\left(\hat{\mu}^{SRI}_\ell-\mu^0_\ell\right)\xrightarrow{\mathcal{L}}N\left(0,\sigma^2_{\ell,SAI}\right)\quad as\quad n_\ell\to\infty,\ \text{ for } \ell=E,R,P,$$

where $\sigma^2_{\ell,SRI} = \mathrm{Var}(e_{\ell,SAIi})$. □

Appendix A.2. Tables: Empirical Type I Error Rates for Scenarios (B) and (C) with Balanced and Unbalanced Designs

Table A1. Empirical Type I error rates for Scenarios (B) and (C) with balanced designs.

		n = 50				n = 100				n = 150			
Scenario	**Case**	**SHT**	**SRI**	**SAI**	**CC**	**SHT**	**SRI**	**SAI**	**CC**	**SHT**	**SRI**	**SAI**	**CC**
(B)	E1	0.041	0.055	0.055	0.068	0.042	0.053	0.059	0.082	0.048	0.058	0.062	0.090
	E2	0.046	0.049	0.050	0.052	0.045	0.051	0.052	0.060	0.042	0.054	0.054	0.066
	E3	0.045	0.045	0.044	0.051	0.037	0.053	0.054	0.060	0.046	0.056	0.058	0.060
	E4	0.050	0.048	0.047	0.052	0.032	0.055	0.058	0.060	0.046	0.052	0.054	0.056
(C)	E1	0.040	0.065	0.064	0.081	0.031	0.052	0.056	0.082	0.046	0.062	0.064	0.082
	E2	0.030	0.052	0.043	0.057	0.046	0.053	0.052	0.063	0.044	0.050	0.056	0.068
	E3	0.034	0.058	0.055	0.060	0.034	0.050	0.053	0.060	0.036	0.048	0.054	0.060
	E4	0.031	0.044	0.043	0.042	0.041	0.052	0.056	0.082	0.048	0.054	0.054	0.075

Note: SHT, SRI, SAI and CC denote Wald-type test approaches based on IPW, regression imputation, AIPW and CC, respectively.

Table A2. Empirical Type I error rates for Scenarios (B) and (C) with unbalanced designs.

			N = 200				N = 300				N = 400			
Scen.	n_E:n_R:n_P	**Case**	**SHT**	**SRI**	**SAI**	**CC**	**SHT**	**SRI**	**SAI**	**CC**	**SHT**	**SRI**	**SAI**	**CC**
(B)	2:1:1	E1	0.035	0.058	0.058	0.068	0.034	0.053	0.059	0.082	0.036	0.062	0.060	0.078
		E2	0.042	0.052	0.046	0.062	0.042	0.053	0.050	0.064	0.046	0.052	0.050	0.070
		E3	0.041	0.066	0.061	0.056	0.038	0.062	0.058	0.042	0.034	0.051	0.049	0.048
		E4	0.033	0.054	0.051	0.050	0.037	0.056	0.051	0.049	0.046	0.052	0.051	0.068
	2:2:1	E1	0.026	0.052	0.050	0.079	0.049	0.068	0.066	0.083	0.030	0.040	0.042	0.080
		E2	0.031	0.051	0.049	0.037	0.041	0.059	0.057	0.047	0.046	0.058	0.056	0.062
		E3	0.026	0.046	0.045	0.040	0.040	0.058	0.052	0.048	0.049	0.056	0.052	0.060
		E4	0.029	0.044	0.042	0.042	0.031	0.046	0.045	0.035	0.030	0.044	0.038	0.044
	3:2:1	E1	0.041	0.060	0.063	0.072	0.037	0.050	0.046	0.069	0.040	0.046	0.046	0.078
		E2	0.032	0.055	0.057	0.054	0.035	0.046	0.043	0.049	0.034	0.044	0.042	0.042
		E3	0.032	0.051	0.045	0.055	0.038	0.056	0.052	0.050	0.032	0.049	0.045	0.046
		E4	0.033	0.047	0.053	0.050	0.032	0.059	0.057	0.052	0.038	0.050	0.048	0.038

Table A2. *Cont.*

Scen.	n_E:n_R:n_P	Case	N = 200				N = 300				N = 400			
			SHT	**SRI**	**SAI**	**CC**	**SHT**	**SRI**	**SAI**	**CC**	**SHT**	**SRI**	**SAI**	**CC**
(C)	2:1:1	E1	0.033	0.067	0.064	0.077	0.029	0.054	0.050	0.069	0.032	0.062	0.054	0.082
		E2	0.043	0.059	0.056	0.051	0.048	0.047	0.047	0.066	0.050	0.058	0.052	0.064
		E3	0.027	0.050	0.049	0.050	0.042	0.062	0.060	0.042	0.028	0.050	0.051	0.059
		E4	0.028	0.051	0.051	0.043	0.026	0.044	0.042	0.042	0.026	0.046	0.045	0.050
	2:2:1	E1	0.021	0.049	0.047	0.081	0.044	0.068	0.067	0.082	0.030	0.042	0.044	0.090
		E2	0.026	0.047	0.043	0.037	0.039	0.058	0.056	0.060	0.024	0.040	0.042	0.054
		E3	0.026	0.046	0.045	0.040	0.058	0.052	0.048	0.049	0.049	0.056	0.052	0.060
		E4	0.029	0.044	0.042	0.042	0.031	0.046	0.045	0.035	0.030	0.044	0.038	0.044
	3:2:1	E1	0.039	0.069	0.069	0.082	0.036	0.051	0.048	0.074	0.034	0.046	0.044	0.082
		E2	0.039	0.058	0.055	0.065	0.037	0.045	0.047	0.052	0.034	0.044	0.042	0.042
		E3	0.031	0.051	0.050	0.046	0.030	0.056	0.054	0.050	0.029	0.050	0.047	0.050
		E4	0.029	0.050	0.052	0.048	0.036	0.058	0.060	0.058	0.031	0.044	0.044	0.045

Note: SHT, SRI, SAI and CC denote Wald-type test approaches based on IPW, terline regression imputation, AIPW and CC, respectively.

Appendix A.3. Figures: Powers for Scenario (A) with a, n, Treatment Effects, Parameters α and γ

Figure A1 presents empirical powers against $a = (\mu_E - \mu_P)/(\mu_R - \mu_P)$ for missingness data mechanism models E1–E4 under the balanced designs, with $n = 50$, 100 and 150. Figure A2 presents empirical powers against $a = (\mu_E - \mu_P)/(\mu_R - \mu_P)$ for missingness data mechanism models E1–E4 under the unbalanced designs with n_E:n_R:n_P = 2:1:1.

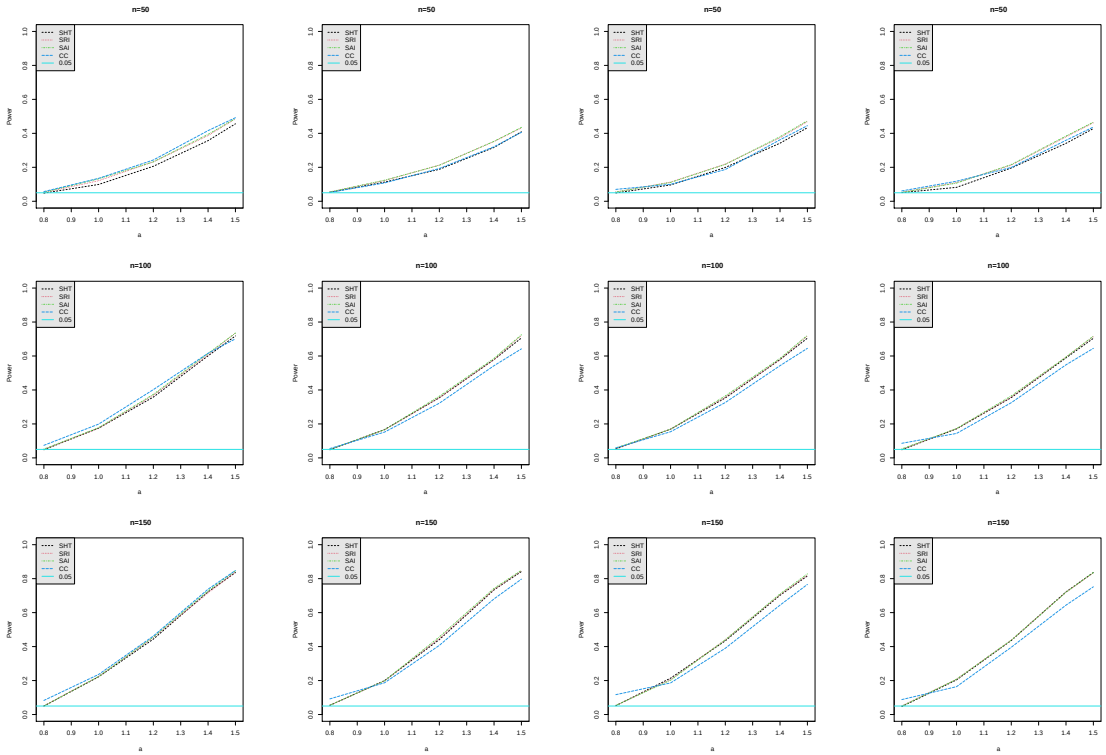

Figure A1. SHT, SRI, SAI and CC represent empirical powers evaluated from IPW, regression imputation, AIPW and CC methods against a for missingness data mechanism model E1 (left panel), E2 (left second panel), E3 (right second panel) and E4 (right panel) for $n = 50$ (the first row), 100 (middle row) and 150 (the last row) under the balanced designs.

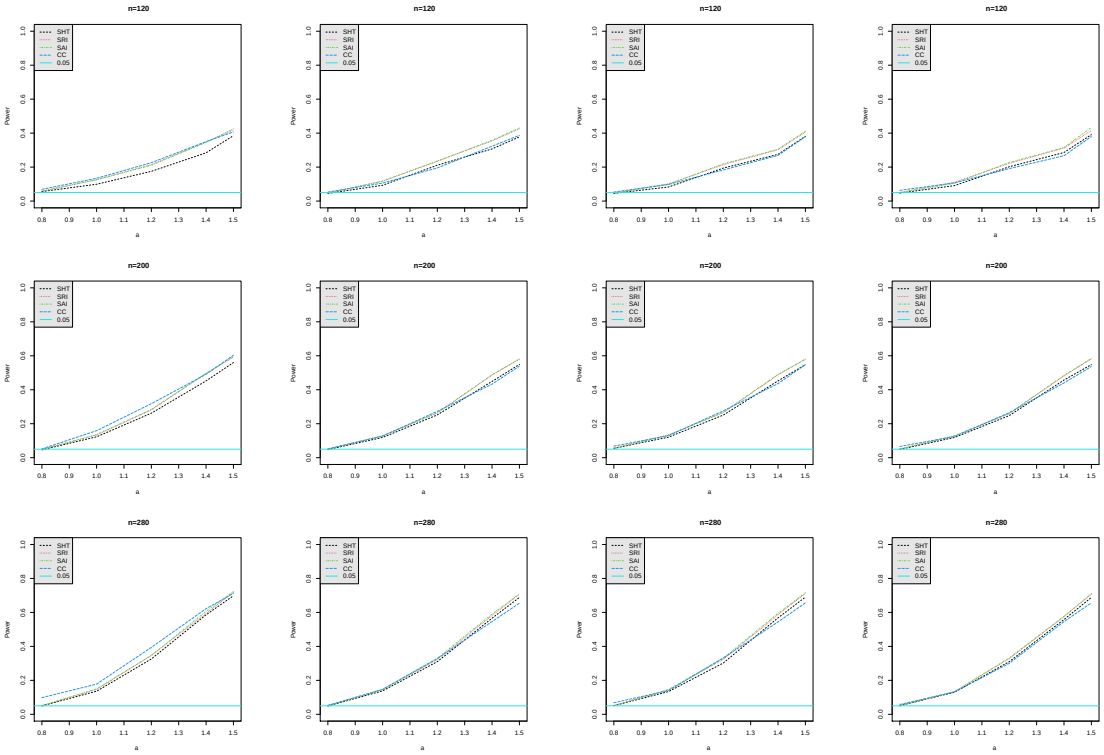

Figure A2. SHT, SRI, SAI and CC represent empirical powers evaluated from IPW, regression imputation, AIPW and CC methods against a for missingness data mechanism models E1 (left panel), E2 (left second panel), E3 (right second panel) and E4 (right panel) for N = 120 (the first row), 200 (middle row) and 280 (the last row) under the unbalanced designs with n_E:n_R:n_P = 2:1:1.

Figure A3 presents empirical powers against the sample size n for missingness data mechanism models E1–E4 under balanced design for $a = 0.4$ and 0.6. Figure A4 presents empirical powers against the sample size n for missingness data mechanism models E1–E4 under unbalanced design with n_E:n_R:n_P = 2:1:1.

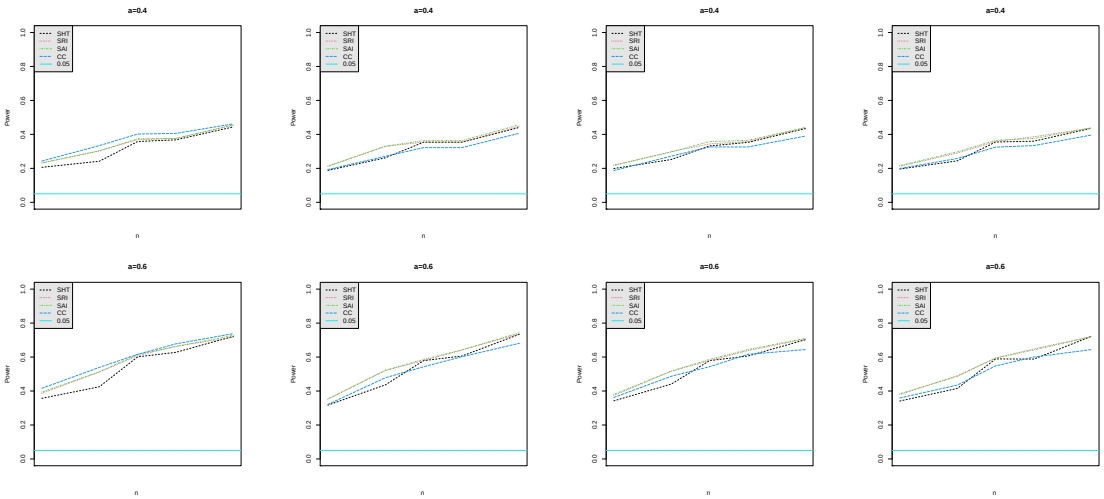

Figure A3. SHT, SRI, SAI and CC represent empirical powers evaluated from IPW, regression imputation, AIPW and CC methods against the sample size n for missingness data mechanism models E1 (left panel), E2 (left second panel), E3 (right second panel) and E4 (right panel) for $a = 0.4$ (upper row) and 0.6 (lower row) under balanced design.

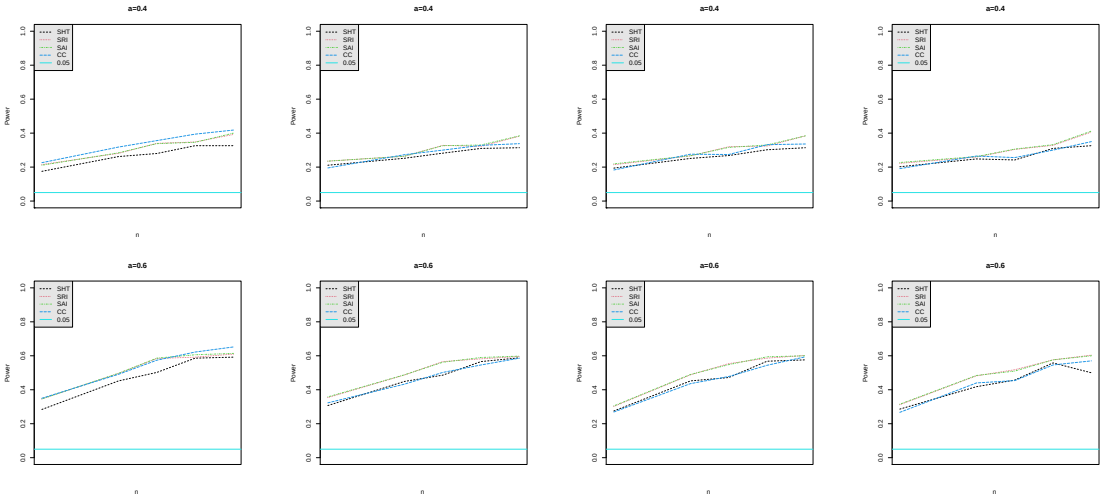

Figure A4. SHT, SRI, SAI and CC represent empirical powers evaluated from IPW, regression imputation, AIPW and CC methods against the sample size n for missingness data mechanism models E1 (left panel), E2 (left second panel), E3 (right second panel) and E4 (right panel) for $a = 0.4$ (upper row) and 0.6 (lower row) under unbalanced design (n_E:n_R:n_P = 2:1:1).

Figure A5 presents empirical powers against treatment effect α_{E1} for four missingness data mechanism models E1–E4 under balanced design. Figure A6 presents empirical powers against treatment effects α_{E1} for four missingness data mechanism models E1–E4 under unbalanced design with n_E:n_R:n_P = 2:1:1.

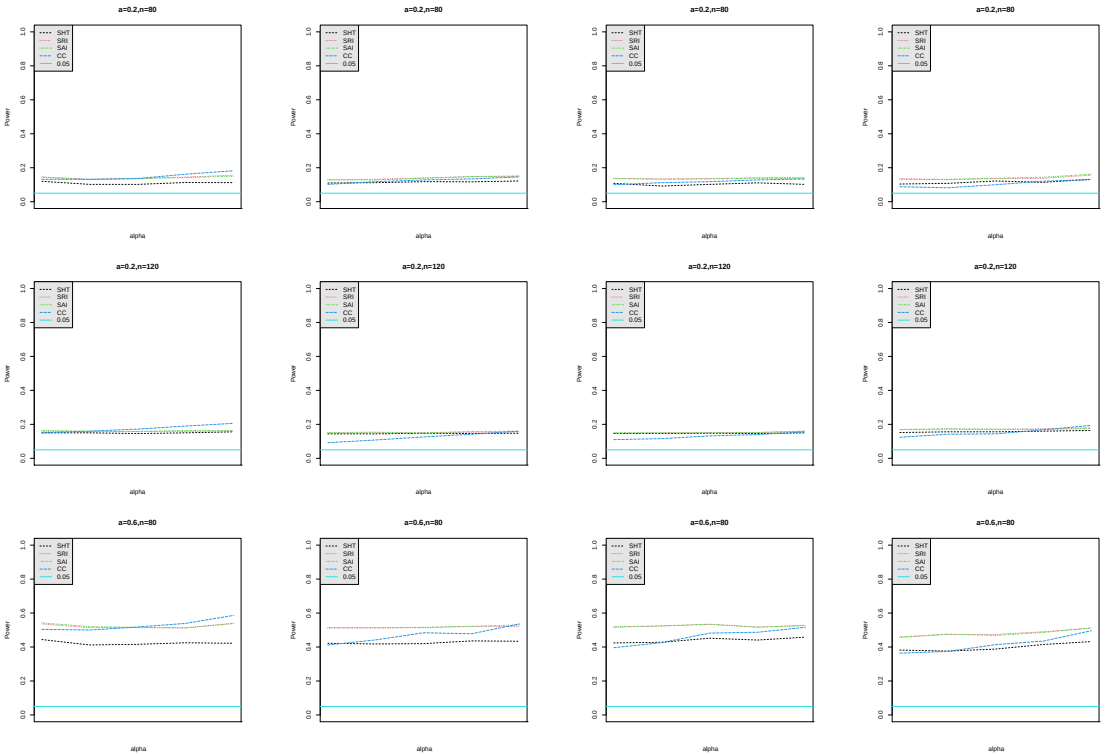

Figure A5. *Cont.*

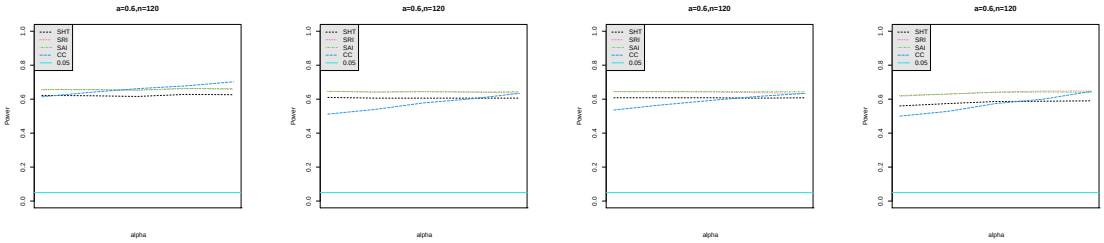

Figure A5. SHT, SRI, SAI and CC represent empirical powers evaluated from IPW, regression imputation, AIPW and CC methods against treatment effect α_{E1} for missingness data mechanism models E1 (left panel), E2 (left second panel), E3 (right second panel) and E4 (right panel) for (a, n) = (0.2,80) (the first row), (0.2,120) (the second row), (0.6,80) (the third row) and (0.6,120) (the last row) under the balanced designs.

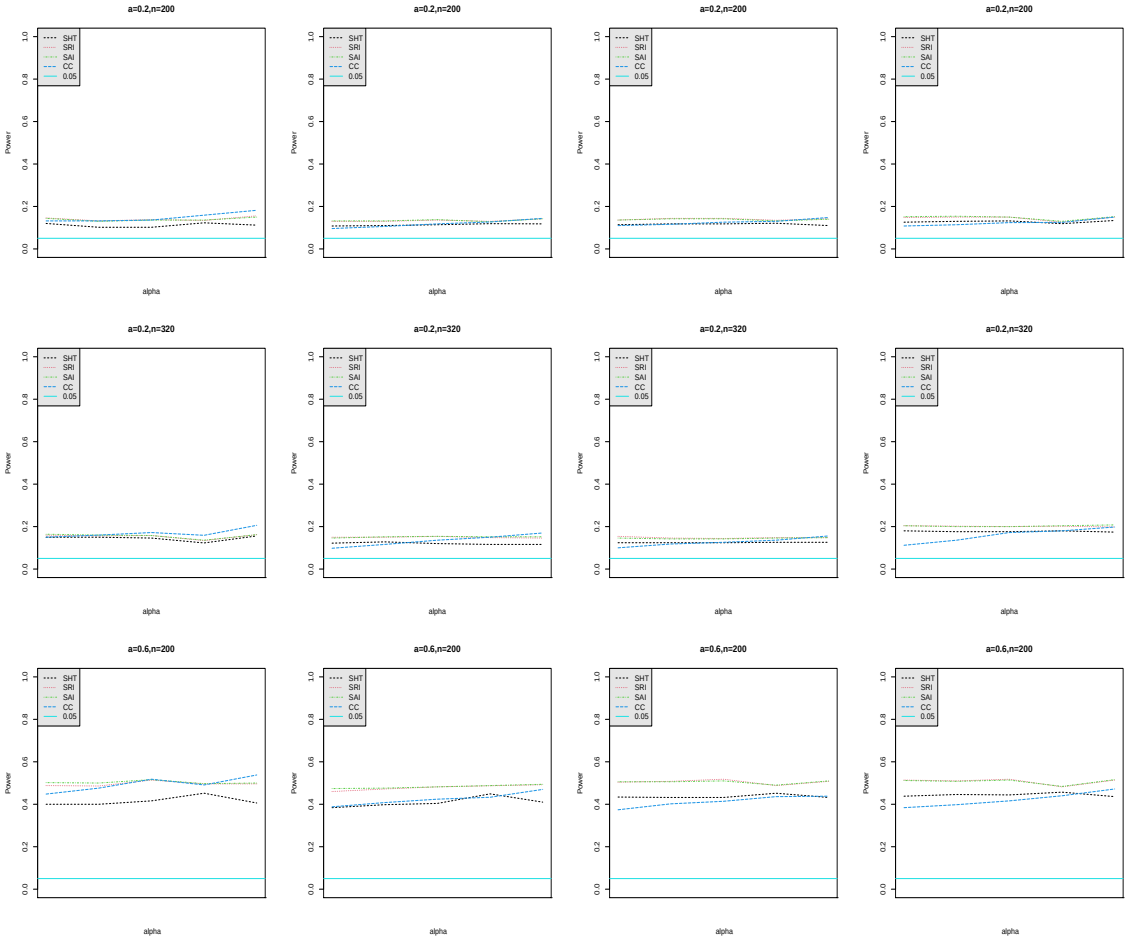

Figure A6. *Cont.*

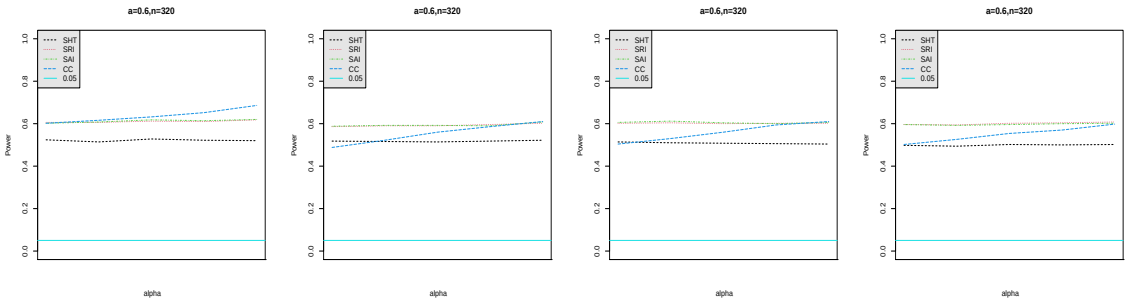

Figure A6. SHT, SRI, SAI and CC represent empirical powers evaluated from IPW, regression imputation, AIPW and CC methods against treatment effect α_{E1} for missingness data mechanism models E1 (left panel), E2 (left second panel), E3 (right second panel) and E4 (right panel) for (a, n) = (0.2,200) (the first row), (0.2,320) (the second row), (0.6,200) (the third row), and (0.6,320) (the last row) under the unbalanced designs with n_E:n_R:n_P = 2:1:1.

Figure A7 presents empirical powers against the tilting parameter γ for three missingness data mechanism models E2–E4 under balanced design. Figure A8 presents empirical powers against the tilting parameter γ for three missingness data mechanism models E2–E4 under unbalanced design with n_E:n_R:n_P = 2:1:1.

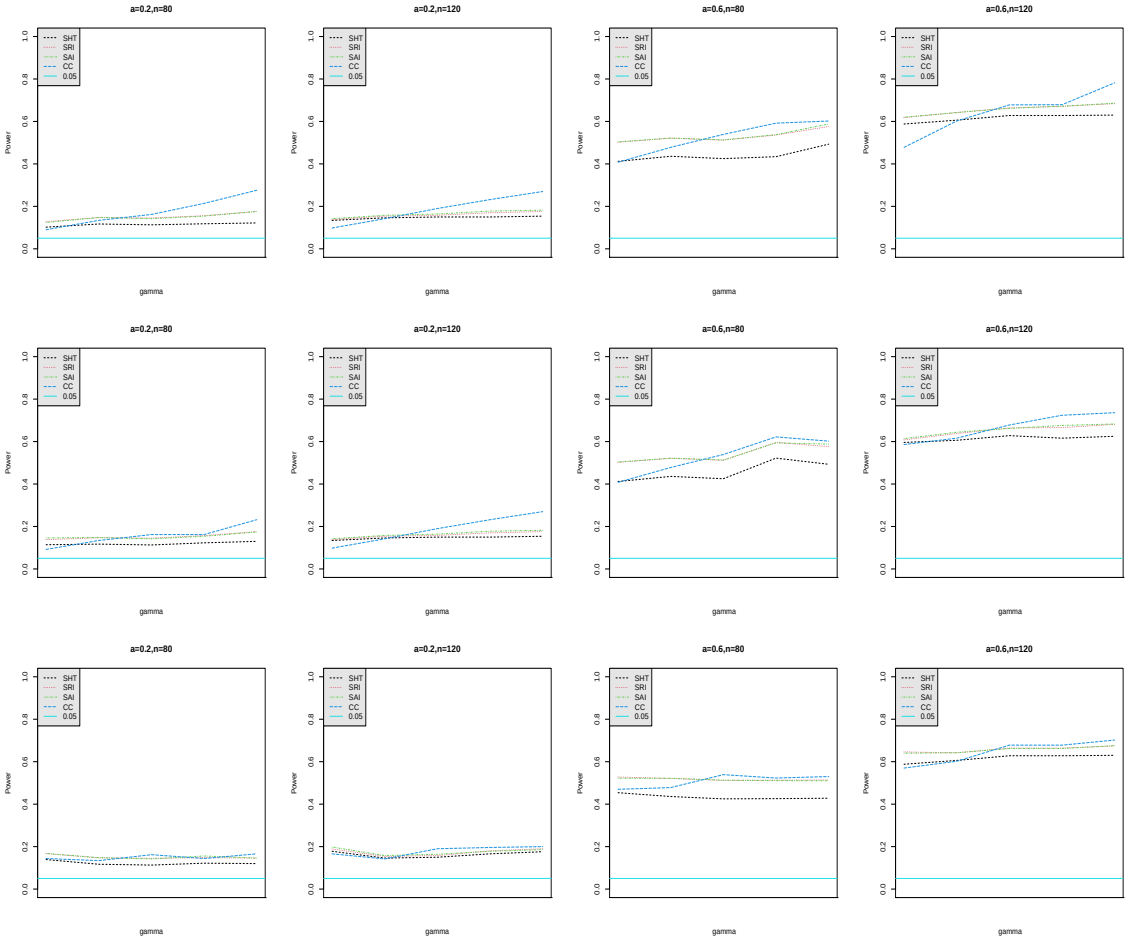

Figure A7. SHT, SRI, SAI and CC represent empirical powers evaluated from IPW, regression imputation, AIPW and CC methods against γ for missingness data mechanism models E2 (the first row), E3 (the middle row) and E4 (the last row) together with (a, n) = (0.2,80), (0.2,120), (0.6,80) and (0.6,120), respectively, under the balanced designs.

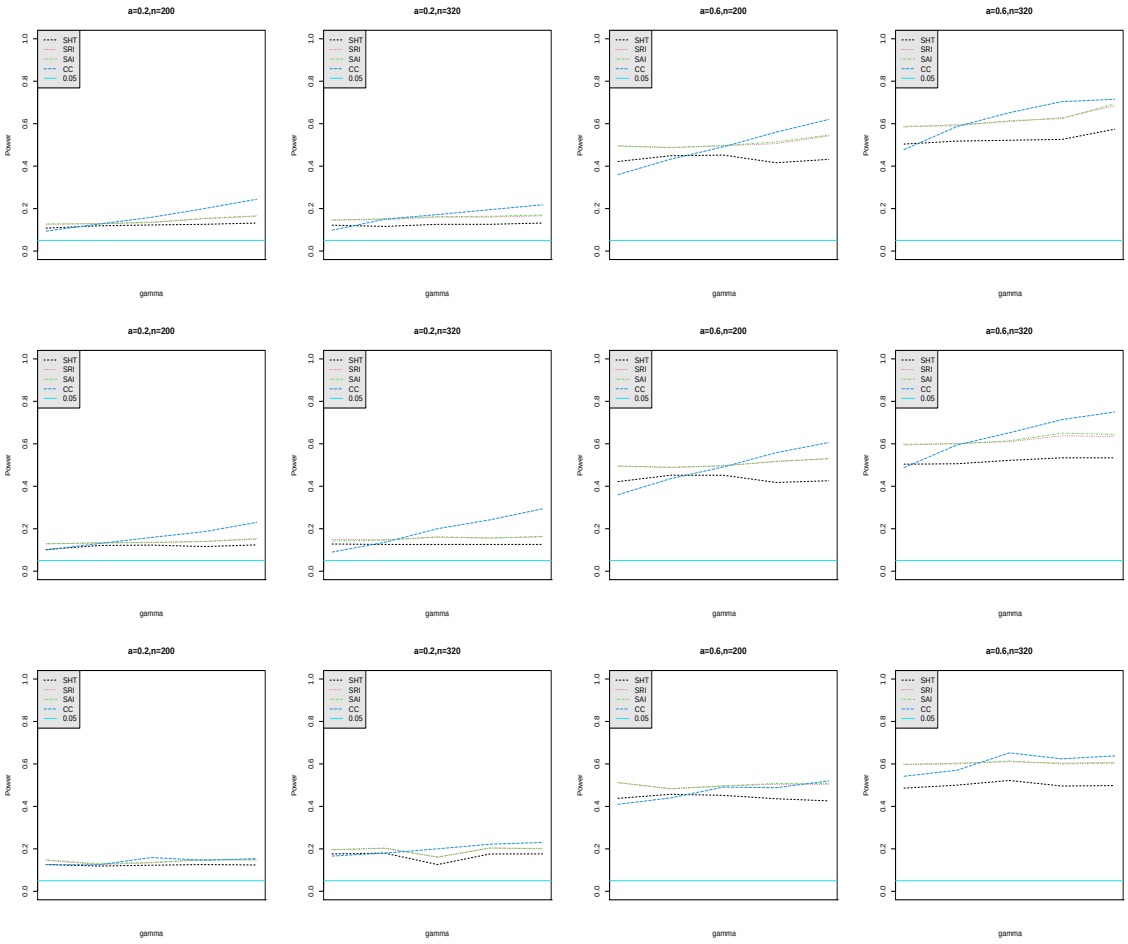

Figure A8. SHT, SRI, SAI and CC represent empirical powers evaluated from IPW, regression imputation, AIPW and CC methods against γ for missingness data mechanism models E2 (the first row), E3 (the middle row) and E4 (the last row) together with (a, n) = (0.2,200), (0.2,320), (0.6,200) and (0.6,320), respectively, under the unbalanced designs with n_E:n_R:n_P = 2:1:1.

References

1. Rothmann, M.D.; Wiens, B.L.; Chan, I.S.F. *Design and Analysis of Non-Inferiority Trials*; (Chapman & Hall/CRC Biostatistics Series, 43); Chapman & Hall/CRC Press: New York, NY, USA, 2011; pp. 149–166.
2. Tang, M.L.; Tang, N.S.; Chan, I.S.; Chan, B.P. Sample size determination for establishing equivalence/noninferiority via ratio of two proportions in matched-pair design. *Biometrics* **2002**, *58*, 957–963. [CrossRef] [PubMed]
3. Tang, N.S.; Tang, M.L.; Chan, S.F. On tests of equivalence via non-unity relative risk for matached-pair design. *Stat. Med.* **2003**, *22*, 1217–1233. [CrossRef] [PubMed]
4. Tang, N.S.; Tang, M.L.; Chan, S.F. Confidence interval construction for proportion difference in small-sample paired studies. *Stat. Med.* **2010**, *24*, 3565–3579. [CrossRef]
5. Wellek, S. Statistical methods for the analysis of two-arm non-inferiority trials with binary outcomes. *Biom. J.* **2005**, *47*, 48–61. [CrossRef] [PubMed]
6. Freitag, G.; Lange, S.; Munk, A. Non-parametric assessment of non-inferiority with censored data. *Stat. Med.* **2006**, *25*, 1201–1217. [CrossRef] [PubMed]
7. Arboretti, G.R.; Bonnini, S.; Solmi, F. Non-parametric two-stage active control testing method for non-inferiority tests. *Quad. Stat.* **2008**, *10*, 73–98.
8. Pesarin, F.; Salmaso, L.; Carrozzo, E.; Arboretti, R. Union-intersection permutation solution for two-sample equivalence testing. *Stat. Comput.* **2016**, *26*, 693–701. [CrossRef]
9. Gamalo, M.A.; Wu, R.; Tiwari, R.C. Bayesian approach to non-inferiority trials for normal means. *Stat. Methods Med. Res.* **2016**, *25*, 221–240. [CrossRef]
10. Tang, M.L.; Tang, N.S. Tests of noninferiority via rate difference for three-arm clinical trials with placebo. *J. Biopharm. Stat.* **2004**, *14*, 337–347. [CrossRef]

11. Koch, G.G.; Röhmel, J. Hypothesis testing in the gold standard design for proving the efficacy of an experimental treatment relative to placebo and a reference. *J. Biopharm. Stat.* **2004**, *14*, 315–325. [CrossRef]
12. Pigeot, I.; Schafer, J.; Rohmel, J.; Hauschke, D. Assessing non-inferiority of a new treatment in a three-arm clinical trial including a placebo. *Stat. Med.* **2003**, *22*, 883–899. [CrossRef] [PubMed]
13. Mielke, M.; Munk, A. The assessment and planning of non-inferiority trials for retention of effect hypotheses-towards a general approach. *arXiv* **2009**, arXiv:0912.4169.
14. Lui, K.; Chang, K. Notes on testing noninferiority in ordinal data under the parallel groups design. *J. Biopharm. Stat.* **2013**, *3*, 1294–1307. [CrossRef] [PubMed]
15. Brannath, W.; Scharpenberg, M.; Schmidt, S. Single-stage, three-arm, adaptive test strategies for non-inferiority trials with an unstable reference. *Stat. Med.* **2022**, *41*, 5033–5045. [CrossRef]
16. Tang, N.S.; Yu, B.; Tang, M.L. Testing non-inferiority of a new treatment in three-arm clinical trials with binary endpoints. *BMC Med. Res. Methodol.* **2014**, *14*, 134. [CrossRef] [PubMed]
17. Tang, N.S.; Yu, B. Simultaneous confidence interval for assessing non-inferiority with assay sensitivity in a three-arm trial with binary endpoints. *Pharm. Stat.* **2020**, *19*, 518–531. [CrossRef] [PubMed]
18. Tang, N.S.; Yu, B. Bayesian sample size determination in a three-arm non-inferiority trial with binary endpoints. *J. Biopharm. Stat.* **2022**, *32*, 768–788. [CrossRef]
19. Paul, E.; Tiwari, R.C.; Chowdhury, S.; Ghosh, S. A more powerful test for three-arm non-inferiority via risk difference: Frequentist and Bayesian approaches. *J. Appl. Stat.* **2022**, *12*, 1–23. [CrossRef]
20. Homma, G.; Diamon, T. Sample size calculation for "gold-standard" noninferiority trials with fixed margins and negative binomial endpoints. *Stat. Biopharm. Res.* **2021**, *13*, 435–447. [CrossRef]
21. Ghosh, S.; Paul, E.; Chowdhury, S.; Tiwari, R.C. New approaches for testing non-inferiority for three-arm trials with Poisson distributed outcomes. *Biostatistics* **2022**, *23*, 136–156. [CrossRef]
22. Ghosh, S.; Guo, W.; Ghosh, S. A hierarchical testing procedure for three arm non-inferiority trials. *Comput. Stat. Data Anal.* **2022**, *174*, 107521. [CrossRef]
23. Scharpenberg, M.; Brannath, W. Simultaneous confidence intervals for an extended Koch-Röhmel design in three-arm non-inferiority trials. *arXiv* **2022**, arXiv:2210.08931.
24. Tang, N.S.; Liang, F. Confidence intervals for assessing non-inferiority with assay sensitivity in a three-arm trial with normally distributed endpoints. *Mathematics* **2022**, *10*, 167. [CrossRef]
25. Munzel, U. Nonparametric non-inferiority analyses in the three-arm design with active control and placebo. *Stat. Med.* **2009**, *28*, 3643–3656. [CrossRef] [PubMed]
26. Tseng, Y.K.; Hsu, K.N. Study design for a three-arm equivalence clinical trial with binomially distributed outcomes. *J. Biopharm. Stat.* **2021**, *31*, 736–744. [CrossRef] [PubMed]
27. Little, R.J.A.; Rubin, D.B. *Statistical Analysis with Missing Data*, 3rd ed.; (Wiley Series in Probability and Statistics); John Wiley & Sons: Hoboken, NJ, USA, 2020; pp. 45–87.
28. Choi, S.C.; Stablein, D.M. Practical tests for comparing two proportions with incomplete data. *J. R. Stat. Soc. Ser. C* **1982**, *31*, 256–262. [CrossRef]
29. Tang, M.L.; Tang, N.S. Exact tests for comparing two paired proportions with incomplete data. *Biom. J.* **2004**, *14*, 72–82. [CrossRef]
30. Maritz, J.S. A permutation paired test allowing for missing values. *Aust. N. Z. J. Stat.* **1995**, *37*, 153–159. [CrossRef]
31. Yu, D.; Lim, J.; Liang, F.; Kim, K.; Kim, B.S.; Jang, W. Permutation test for incomplete paired data with application to cDNA microarray data. *Comput. Stat. Data Anal.* **2013**, *56*, 510–521. [CrossRef]
32. Pesarin, F. *Multivariate Permutation Tests: With Applications in Biostatistics*; (Wiley Series in Probability and Statistics); John Wiley & Sons: Chichester, UK, 2001; pp. 279–303.
33. Pesarin, F.; Celant, G.; Salmaso, L. Two sample permutation tests for repeated measures with missing values. *J. Appl. Stat. Sci.* **2000**, *9*, 291–304.
34. Robins, J.; Rotnitzky, A.; Zhao, L.P. Estimation of regression coefficients when some regressors are not always observed. *J. Am. Stat. Assoc.* **1994**, *89*, 846–866. [CrossRef]
35. Lee, S.Y.; Tang, N.S. Bayesian analysis of nonlinear structural equation models with nonignorable missing data. *Psychometrika* **2006**, *71*, 541–564. [CrossRef]
36. Wang, Z.Q.; Tang, N.S. Bayesian quantile regression with mixed discrete and nonignorable missing covariates. *Bayesian Anal.* **2020**, **15**, 579–604. [CrossRef]
37. Kim, J.K.; Yu, C.L. A semiparametric estimation of mean functionals with nonignorable missing data. *J. Am. Stat. Assoc.* **2011**, *106*, 157–165. [CrossRef]
38. Tang, N.S.; Zhao, P.Y.; Zhu, H.T. Empirical likelihood for estimating equations with nonignorable missing data. *Stat. Sin.* **2014**, *24*, 723–747. [PubMed]
39. Qin, J.; Lawless, J. Empirical likelihood and general estimating equations. *Ann. Stat.* **1994**, *22*, 300–325. [CrossRef]
40. Choi, S.C.; Stablein, D.M. Comparing incomplete paired binomial data under non-random mechanisms. *Stat. Med.* **1988**, *7*, 929–939. [CrossRef]
41. Li, H.Q.; Tian, G.L.; Tang, N.S.; Cao, H.Y. Assessing non-inferiority for incomplete paired-data under non-ignorable missing mechanism. *Comput. Stat. Data Anal.* **2018**, *127*, 69–81. [CrossRef]

42. Hida, E.; Tango, T. On the three-arm non-inferiority trial including a placebo with a prespecified margin. *Stat. Med.* **2011**, *30*, 224–231. [CrossRef]
43. Kieser, M.; Friede, T. Planning and analysis of three-arm non-inferiority trials with binary endpoints. *Stat. Med.* **2007**, *26*, 253–273. [CrossRef]
44. Ghosh, P.; Nathoo, F.; Gönen, M.; Tiwari, R.C. Assessing non-inferiority in a three-arm trial using the Bayesian approach. *Stat. Med.* **2011**, *30*, 1795–1808. [CrossRef] [PubMed]
45. Morikawa, K.; Kim, J.K.; Kano, Y. Semiparametric maximum likelihood estimation with data missing not at random. *Can. J. Stat.* **2017**, *45*, 393–409. [CrossRef]
46. Arnastauskaite, J.; Ruzgas, T.; Brazenas, M. A new goodness of fit test for multivariate normality and comparative simulation study. *Mathematics* **2021**, *9*, 3003. [CrossRef]
47. Elbouch, S.E.; Michel, O.J.; Comon, P. A normality test for multivariate dependent samples. *Signal Process.* **2022**, *201*, 108705. [CrossRef]
48. Cook, T.; Zea, R. Missing data and sensitivity analysis for binary data with implications for sample size and power of randomized clinical trials. *Stat. Med.* **2019**, *39*, 192–204. [CrossRef] [PubMed]
49. Zhao, P.Y.; Tang, M.L.; Tang, N.S. Robust estimation of distribution functions and quantiles with non-ignorable missing data. *Can. J. Stat.* **2013**, *129*, 193–205. [CrossRef]