Journal of the American Statistical Association

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/uasa20

To Adjust or not to Adjust? Estimating the Average Treatment Effect in Randomized Experiments with Missing Covariates

Anqi Zhao & Peng Ding

To cite this article: Anqi Zhao & Peng Ding (2022): To Adjust or not to Adjust? Estimating the Average Treatment Effect in Randomized Experiments with Missing Covariates, Journal of the American Statistical Association, DOI: 10.1080/01621459.2022.2123814

To link to this article: https://doi.org/10.1080/01621459.2022.2123814

JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION
2022, VOL. 00, NO. 0, 1–11: Theory and Methods
https://doi.org/10.1080/01621459.2022.2123814

Taylor & Francis
Taylor & Francis Group

To Adjust or not to Adjust? Estimating the Average Treatment Effect in Randomized Experiments with Missing Covariates

Anqi Zhao[a] and Peng Ding[b]

[a]Department of Statistics and Data Science, National University of Singapore, Singapore; [b]Department of Statistics, University of California, Berkeley, CA

ABSTRACT

Randomized experiments allow for consistent estimation of the average treatment effect based on the difference in mean outcomes without strong modeling assumptions. Appropriate use of pretreatment covariates can further improve the estimation efficiency. Missingness in covariates is nevertheless common in practice, and raises an important question: should we adjust for covariates subject to missingness, and if so, how? The unadjusted difference in means is always unbiased. The complete-covariate analysis adjusts for all completely observed covariates, and is asymptotically more efficient than the difference in means if at least one completely observed covariate is predictive of the outcome. Then what is the additional gain of adjusting for covariates subject to missingness? To reconcile the conflicting recommendations in the literature, we analyze and compare five strategies for handling missing covariates in randomized experiments under the design-based framework, and recommend the missingness-indicator method, as a known but not so popular strategy in the literature, due to its multiple advantages. First, it removes the dependence of the regression-adjusted estimators on the imputed values for the missing covariates. Second, it does not require modeling the missingness mechanism, and yields consistent estimators even when the missingness mechanism is related to the missing covariates and unobservable potential outcomes. Third, it ensures large-sample efficiency over the complete-covariate analysis and the analysis based on only the imputed covariates. Lastly, it is easy to implement via least squares. We also propose modifications to it based on asymptotic and finite sample considerations. Importantly, our theory views randomization as the basis for inference, and does not impose any modeling assumptions on the data-generating process or missingness mechanism. Supplementary materials for this article are available online.

ARTICLE HISTORY
Received November 2021
Accepted August 2022

KEYWORDS
Efficiency; Imputation; Missingness pattern; Randomized controlled trial; Regression adjustment; Robust standard error

1. Introduction

Randomized experiments are the gold standard for estimating treatment effects, and justify simple comparisons of mean outcomes across treatment groups (Neyman 1923). Appropriate adjustment for pretreatment covariates further promises additional gains in estimation efficiency (Fisher 1935; Lin 2013). In particular, Lin (2013) showed that the coefficient of the treatment from the ordinary least squares (OLS) fit of the outcome on the treatment, centered covariates, and their interactions is a consistent and asymptotically efficient estimator for the average treatment effect, and established the associated Eicker–Huber–White robust standard error as a convenient approximation to the true standard error. His theory holds even when the linear model is misspecified.

Missingness in covariates is ubiquitous in experiments in biomedical and social sciences, and imposes a dilemma on subsequent inference. We can simply ignore the covariates and proceed with the unadjusted difference in means, which is unbiased and consistent under complete randomization. An immediate improvement is the *complete-covariate analysis*, which adjusts for only the covariates that are observed for all units. It is asymptotically more efficient than the unadjusted estimator if at least one completely observed covariate is prognostic to the outcome. Alert to the waste of information under complete-covariate analysis by discarding all information in the incomplete covariates, alternative approaches like maximum likelihood methods, Bayesian methods, and multiple imputation make use of all available covariates and are likely to further improve efficiency (Rubin 1987; Little 1992; Ibrahim et al. 2005; Little and Rubin 2019). These more sophisticated methods nevertheless rely on additional assumptions on the outcome model or missingness mechanism, and do not strictly dominate the simpler unadjusted difference in means and complete-covariate analysis when the models are misspecified (White and Carlin 2010). Then a natural question arises: should we adjust for the missing covariates or not?

Our answer to the above question is yes. We propose to impute the missing covariates with zeros, augment the imputed covariates with the missingness indicators, and then apply Lin's (2013) procedure for regression adjustment. This method becomes intuitive if the missingness is not affected by the treatment such that we can view the missingness indicators as a set of fully observed pretreatment covariates. The resulting

CONTACT Peng Ding *pengdingpku@berkeley.edu* Department of Statistics, University of California, Berkeley, CA.
Supplementary materials for this article are available online. Please go to *www.tandfonline.com/r/JASA*.

missingness-indicator method has multiple advantages. First, it is invariant to the imputed values for the missing covariates, and allows for the convenient choice of imputing by zeros with no loss of generality. Second, it does not require modeling of the missingness mechanism, and remains consistent for the average treatment effect even when the missingness mechanism depends on the missing covariates and unobservable potential outcomes, a scenario analogous to missing not at random under the superpopulation framework (Rubin 1976; Little and Rubin 2019). Third, it is asymptotically more efficient than the complete-covariate analysis and *single imputation*. Lastly, it can be easily implemented via standard software packages for OLS.

This missingness-indicator method has been used before. Cohen and Cohen (1975, chap. 7) proposed to use it in regression analysis. Rosenbaum and Rubin (1984, Appendix B) suggested it in matching for causal inference with observational studies. Gerber and Green (2012, sec. 7.7) recommended it for covariate adjustment in randomized experiments. See D'Agostino and Rubin (2000), Mattei (2009), Rosenbaum (2010, secs. 9.4 and 13.4), Fogarty et al. (2016), and Chong et al. (2016) for applications of this method to obtain observational studies and randomized experiments. See Anderson, Basilevsky, and Hum (1983) and Groenwold et al. (2012) for a review. In contrast, Greenland and Finkle (1995) and Donders et al. (2006) criticized this method and argued against its use for observational studies. Specifically, Greenland and Finkle (1995) evaluated this method in the context of logistic regression for observational studies and reported severe bias even when the data are missing completely at random; Donders et al. (2006) offered a similar discussion. Miettinen (1985) acknowledged its ease of use, but also pointed out its limitation of representing only partial control when applied to confounders.

Our recommendation does not contradict these criticisms in the literature. The fundamental difference between randomized experiments and observational studies explains the seemingly contradictory recommendations. In particular, randomization balances the pretreatment covariates along with missingness indicators on average across treatment groups. Using them in Lin (2013)'s procedure ensures consistent estimation of the average treatment effect regardless of the missingness mechanism. Our results echo White and Thompson (2005) and Carpenter and Kenward (2007) without assuming that the covariates and outcomes are jointly normal. Rather, we evaluate different methods under the design-based framework, also known as the randomization-based framework, free of any modeling assumptions (Neyman 1923; Imbens and Rubin 2015). The above theoretical guarantees of the missingness-indicator method thus hold even when the regression models are misspecified. This is a key distinction between our results and those for correctly specified regression models with missing covariates (Rubin 1987; Little 1992; Robins, Rotnitzky, and Zhao 1994; Jones 1996; Ibrahim et al. 2005). In contrast, causal inference is fundamentally more challenging in observational studies with missing covariates. Rosenbaum and Rubin (1984) showed the identifiability of the average treatment effect under the unconfoundedness assumption given the observed covariates and missingness pattern. Under this assumption, the missingness-indicator method can be used to estimate the average treatment effect if the linear model is correct. As pointed out by Yang, Wang, and Ding (2019), however, this assumption is scientifically implausible by requiring the set of confounders to depend on the missingness pattern. Without this assumption, the average treatment effect is not identifiable. See Ding and Geng (2014) and Yang, Wang, and Ding (2019) for alternative strategies.

Moreover, we propose a modification to the missingness-indicator method that captures additional efficiency in large samples. Specifically, the *missingness-pattern method* stratifies the data based on the missingness patterns, and applies Lin (2013)'s approach based on the available covariates within each stratum. It is closely related to post-stratification (Miratrix, Sekhon, and Yu 2013) if we view the type of missingness pattern as a discrete covariate, and can be seen as an extension of the *available-covariate analysis* proposed by Wilks (1932), Matthai (1951), Glasser (1964), and Haitovsky (1968). The resulting estimator allows for heterogeneous adjustments across different missingness patterns, and thereby promises additional asymptotic efficiency over the missingness-indicator method if the covariates and missingness patterns affect the treatment effects in nonadditive ways. We recommend this method if the sample sizes within all missingness patterns are large enough to justify the application of Lin (2013)'s estimator.

We use the following notation. For a finite population $\{(u_i, v_i) : i \in \mathcal{I}\}$, let $\bar{u} = |\mathcal{I}|^{-1} \sum_{i\in\mathcal{I}} u_i$, $\bar{v} = |\mathcal{I}|^{-1} \sum_{i\in\mathcal{I}} v_i$, and $S_{uv} = (|\mathcal{I}| - 1)^{-1} \sum_{i\in\mathcal{I}} (u_i - \bar{u})(v_i - \bar{v})^{\mathrm{T}}$ denote the finite population means and covariance, respectively. We specify the composition of $\mathcal{I}$ in the context. In the case of $u_i = v_i$, we also occasionally write S_{uu} as S_u^2. For $u_i \in \mathbb{R}$ and $v_i \in \mathbb{R}^p$, let $u_i \sim v_i$ denote the OLS fit of u_i on v_i over $i \in \mathcal{I}$. Importantly, we do not invoke the stochastic assumptions for the underlying linear models, but evaluate the sampling properties of the numeric outputs of OLS from the design-based perspective. We focus on the robust standard errors from OLS because the classic standard errors do not have the desired design-based properties even in simpler cases (Freedman 2008; Lin 2013).

2. Setting

2.1. Regression Estimators when all Covariates are Observed

Consider an intervention of two levels, $z = 0, 1$, and a finite population of N units, $i = 1, \ldots, N$. Let $Y_i(z)$ be the potential outcome of unit i under treatment level z. The individual treatment effect is $\tau_i = Y_i(1) - Y_i(0)$, and the finite population average treatment effect is $\tau = N^{-1} \sum_{i=1}^N \tau_i = \bar{Y}(1) - \bar{Y}(0)$, where $\bar{Y}(z) = N^{-1} \sum_{i=1}^N Y_i(z)$.

We focus on complete randomization, which assigns completely at random N_z units to treatment level z with $N_0 + N_1 = N$. Let $e_z = N_z/N$ for $z = 0, 1$. Let $Z_i \in \{0, 1\}$ denote the treatment assignment of unit i. The observed outcome is $Y_i = Z_i Y_i(1) + (1 - Z_i) Y_i(0)$.

Let $\widehat{Y}(z) = N_z^{-1} \sum_{i:Z_i=z} Y_i$ be the average observed outcome under treatment level z. The difference in means $\widehat{\tau}_{\mathrm{N}} = \widehat{Y}(1) - \widehat{Y}(0)$ is unbiased for τ, and equals the coefficient of Z_i from the OLS fit of the unadjusted regression $Y_i \sim 1 + Z_i$ over $i = 1, \ldots, N$. We use the subscript "N" to signify Neyman (1923).

The presence of covariates promises the opportunity to improve estimation efficiency. Let $x_i = (x_{i1}, \ldots, x_{iJ})^{\mathrm{T}}$ be the J-dimensional covariate vector for unit i. Fisher (1935) suggested an estimator $\widehat{\tau}_{\mathrm{F}}$ for τ, which equals the coefficient of Z_i from the OLS fit of the additive regression $Y_i \sim 1 + Z_i + x_i$ over $i = 1, \ldots, N$. Lin (2013) recommended an improved estimator, $\widehat{\tau}_{\mathrm{L}}$, as the coefficient of Z_i from the OLS fit of the fully interacted regression $Y_i \sim 1 + Z_i + (x_i - \bar{x}) + Z_i(x_i - \bar{x})$ over $i = 1, \ldots, N$ with centered covariates and treatment-covariates interactions, and showed its asymptotic efficiency over $\widehat{\tau}_{\mathrm{N}}$ and $\widehat{\tau}_{\mathrm{F}}$. We summarize the results in Lemma 1 below, with the subscripts N, F, and L signifying quantities associated with the unadjusted, additive, and fully interacted regressions, respectively. We adopt the finite population design-based framework, which conditions on the potential outcomes and covariates and views $(Z_i)_{i=1}^N$ as the sole source of randomness when evaluating estimators.

Let $\gamma_{\mathrm{L},z} = (S_x^2)^{-1} S_{xY(z)}$ denote the coefficient vector of x_i from the OLS fit of $Y_i(z) \sim 1 + x_i$ over $i = 1, \ldots, N$, where S_x^2 is the finite population covariance of the x_is, and $S_{xY(z)}$ is the finite population covariance of the x_is and $Y_i(z)$s, respectively. Let $\gamma_{\mathrm{F}} = e_0\gamma_{\mathrm{L},0} + e_1\gamma_{\mathrm{L},1}$, and let $S_{z,*}^2$ be the finite population variance of $\{Y_{*,i}(z)\}_{i=1}^N$ for $z = 0, 1$ and $* = \mathrm{N}, \mathrm{F}, \mathrm{L}$, with $Y_{\mathrm{N},i}(z) = Y_i(z)$, $Y_{\mathrm{F},i}(z) = Y_i(z) - x_i^{\mathrm{T}}\gamma_{\mathrm{F}}$, and $Y_{\mathrm{L},i}(z) = Y_i(z) - x_i^{\mathrm{T}}\gamma_{\mathrm{L},z}$. Let $S_{\tau,*}^2$ be the finite population variance of $(\tau_{*,i})_{i=1}^N$, where $\tau_{*,i} = Y_{*,i}(1) - Y_{*,i}(0)$. To simplify the presentation, we relegate the regularity conditions to Condition S4 in the supplementary materials. In particular, Condition S4 ensures that e_z, $S_{z,*}^2$, and $S_{\tau,*}^2$ have finite limits for all $z = 0, 1$ and $* = \mathrm{N}, \mathrm{F}, \mathrm{L}$. We will also use the same symbols to denote their respective limits when no confusion would arise.

Lemma 1. Assume complete randomization and Condition S4. Then $\sqrt{N}(\widehat{\tau}_* - \tau) \rightsquigarrow \mathcal{N}(0, v_*)$ for $* = \mathrm{N}, \mathrm{F}, \mathrm{L}$ with $v_* = e_0^{-1}S_{0,*}^2 + e_1^{-1}S_{1,*}^2 - S_{\tau,*}^2$ and $v_{\mathrm{L}} \leq v_{\mathrm{N}}, v_{\mathrm{F}}$. Further let $\widehat{\mathrm{se}}_*$ be the robust standard error of $\widehat{\tau}_*$ from the corresponding OLS fit. Then $N\widehat{\mathrm{se}}_*^2 - v_* = S_{\tau,*}^2 + o_{\mathrm{P}}(1)$ with $S_{\tau,*}^2 \geq 0$.

Lemma 1 reviews the classic results from Neyman (1923), Freedman (2008), and Lin (2013). The unadjusted, additive, and fully interacted regressions yield consistent and asymptotically normal estimators for τ, with the corresponding robust standard errors being asymptotically conservative for estimating the true standard errors. This justifies the large-sample Wald-type inference based on $(\widehat{\tau}_*, \widehat{\mathrm{se}}_*)$ for $* = \mathrm{N}, \mathrm{F}, \mathrm{L}$. Asymptotically, $\widehat{\tau}_{\mathrm{L}}$ is the most efficient, whereas $\widehat{\tau}_{\mathrm{F}}$ may be more efficient or less efficient than the difference in means. In particular, Lin (2013) showed that $v_{\mathrm{F}} = v_{\mathrm{L}}$ if $e_0 = e_1$. This ensures that $\widehat{\tau}_{\mathrm{F}}$ is asymptotically as efficient as $\widehat{\tau}_{\mathrm{L}}$ when the treatment groups are of equal sizes.

Remark 1. Let $\widehat{\tau}_x = \widehat{x}(1) - \widehat{x}(0)$ denote the difference in covariate means, with $\widehat{x}(z) = N_z^{-1}\sum_{i:Z_i=z} x_i$. Li and Ding (2020) showed that $v_{\mathrm{L}} = v_{\mathrm{N}}(1 - R^2)$, where R^2 is the squared multiple correlation between $\widehat{\tau}_{\mathrm{N}}$ and $\widehat{\tau}_x$. Including more covariates in Lin (2013)'s specification can thus never harm the asymptotic efficiency. We give a more general result in Lemma S6 in the supplementary materials. An important caveat is that increased model complexity can increase variance of the estimator in finite samples. Including more covariates in Lin (2013)'s specification thus ensures asymptotic efficiency at the cost of finite sample performance. The same argument extends to the choice between additive and fully interacted specifications for regression adjustment. The inclusion of interactions effectively doubles the model complexity, such that the asymptotically less efficient $\widehat{\tau}_{\mathrm{F}}$ can have better performance in finite samples.

Remark 2. The asymptotic conservativeness of $\widehat{\mathrm{se}}_{\mathrm{L}}$ is specific to finite population inference (Neyman 1923; Lin 2013; Imbens and Rubin 2015; Li and Ding 2020), where we condition on the covariates for all analyses. Superpopulation inference, in contrast, assumes that x_i's are independent and identically distributed samples from some superpopulation. The resulting $\widehat{\tau}_{\mathrm{L}}$ will have extra variability due to the centering of the covariates; see Berk et al. (2013), Negi and Wooldridge (2021), Zhao and Ding (2021), and Ye et al. (2022).

2.2. Missing Data and Running Examples

The construction of $\widehat{\tau}_*$ $(* = \mathrm{F}, \mathrm{L})$ assumes that the covariates are fully observed for all N units. A key question is how to adapt when some covariates are only partially available.

Let $M_i = (M_{i1}, \ldots, M_{iJ})^{\mathrm{T}} \in \{0, 1\}^J$ be the missingness indicators for unit i with $M_{ij} = 1$ if x_{ij} is missing and $M_{ij} = 0$ if otherwise. The possible values of M_i define 2^J possible missingness patterns, indexed by $m = (m_1, \ldots, m_J)^{\mathrm{T}} \in \{0, 1\}^J$. Not all 2^J missingness patterns need to be present in any given dataset. Let $N_{(m)} = N_{(m_1,\ldots,m_J)} = \sum_{i=1}^N 1(M_i = m)$ be the number of units with missingness pattern $m = (m_1, \ldots, m_J)^{\mathrm{T}}$, and let $\mathcal{M} = \{m : N_{(m)} > 0\}$ be the set of missingness patterns that are present in the dataset. We use the following examples to illustrate these notation.

Example 1. Consider the case with $J = 1$ covariate, $x_i = x_{i1}$, for $i = 1, \ldots, N$. The missingness indicators are $M_i = M_{i1} \in \{0, 1\}$, and suggest two possible missingness patterns, $m \in \{0, 1\}$, with $N_{(0)} = \sum_{i=1}^N 1(M_i = 0)$ and $N_{(1)} = \sum_{i=1}^N 1(M_i = 1)$:

Missingness pattern (M_i)	x_i	Number of units
0	observed	$N_{(0)}$
1	missing	$N_{(1)}$

Example 2. Consider the case with $J = 2$ covariates, $x_i = (x_{i1}, x_{i2})^{\mathrm{T}}$, for $i = 1, \ldots, N$. The missingness indicators are $M_i = (M_{i1}, M_{i2})^{\mathrm{T}} \in \{0, 1\}^2$, and suggest $2^2 = 4$ possible missingness patterns, $m = (m_1, m_2)^{\mathrm{T}} \in \{0, 1\}^2$, with $N_{(m)} = \sum_{i=1}^N 1(M_{i1} = m_1, M_{i2} = m_2)$:

Missingness pattern (M_i)	x_{i1}	x_{i2}	Number of units
$(0, 0)^{\mathrm{T}}$	observed	observed	$N_{(0,0)}$
$(0, 1)^{\mathrm{T}}$	observed	missing	$N_{(0,1)}$
$(1, 0)^{\mathrm{T}}$	missing	observed	$N_{(1,0)}$
$(1, 1)^{\mathrm{T}}$	missing	missing	$N_{(1,1)}$

Depending on the specific application under consideration, the missingness may or may not depend on the treatment assignment. To simplify the presentation, we focus on the case

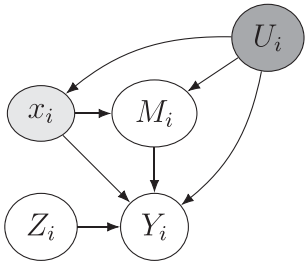

Figure 1. A direct acyclic graph illustrating Condition 1. White nodes represent the treatment assignment Z_i, missingness indicators M_i, and outcome Y_i that are fully observed. The light gray node represents the covariate vector x_i with missing values. The dark node represents the unmeasured covariate U_i.

where missingness is unaffected by the treatment assignment in the main paper, and discuss the case of assignment-dependent missingness in the supplementary materials. Let $M_i(z) = (M_{i1}(z), \ldots, M_{iJ}(z))^{\mathrm{T}}$ be the potential value of M_i if unit i were assigned to treatment level z. Condition 1 formalizes the notion of treatment-independent missingness in terms of $M_i(z)$'s.

Condition 1. $M_i(0) = M_i(1) = M_i$ for all $i = 1, \ldots, N$.

Condition 1 ensures that the missingness is unaffected by the treatment assignment such that the M_i's are effectively a set of fully observed pretreatment covariates. If the missingness happens before the treatment assignment, it cannot be affected by the treatment and Condition 1 holds automatically. If the covariates are collected retrospectively after the experiment, the missingness may be affected by the treatment and Condition 1 may be violated. The former case is arguably more common in randomized experiments (White and Thompson 2005; Carpenter and Kenward 2007; Sullivan et al. 2018). This suggests the mildness of Condition 1. Figure 1 is a graphical model, illustrating the generality of Condition 1. Assume that U_i is an unmeasured covariate for $i = 1, \ldots, N$. Graphically, there is no link between Z_i and M_i, encoding Condition 1. Condition 1 allows for unmeasured common causes of x_i, M_i, and Y_i, such that the missingness can depend on both the missing covariates and unobserved potential outcomes.

3. Five Strategies for Handling Missing Covariates

We outline in this section five strategies for handling missing covariates. The simplest options are the *complete-case analysis* that uses only units with all covariates observed and the *complete-covariate analysis* that uses only covariates that are completely observed for all units. *Single imputation* instead first fills in the missing covariates with some values and then proceeds with the standard complete-data analysis with the imputed data. When the missingness indicators act as pretreatment covariates, it is also natural to include them directly in OLS, motivating the *missingness-indicator method*. The *missingness-pattern method* stratifies the data based on the missingness patterns, and performs one separate analysis for each missingness pattern.

3.1. Complete-Case Analysis

Most standard software routines adopt the *complete-case analysis* as default by dropping all units with any missingness in covariates. Let $C_i = 1(M_i = 0_J)$ be the complete-case indicator for unit i, with $C_i = 1$ if and only if x_i is fully observed. The complete-case analysis proceeds with only the $N^{\mathrm{cc}} = \sum_{i=1}^{N} C_i$ complete cases, $\{i : C_i = 1\}$, and fits

$$Y_i \sim 1 + Z_i + x_i, \tag{1}$$
$$Y_i \sim 1 + Z_i + (x_i - \bar{x}^{\mathrm{cc}}) + Z_i(x_i - \bar{x}^{\mathrm{cc}}) \tag{2}$$

over $\{i : C_i = 1\}$ under the additive and fully interacted specifications, respectively, where $\bar{x}^{\mathrm{cc}} = (N^{\mathrm{cc}})^{-1} \sum_{i:C_i=1} x_i$. We can then use the coefficients of Z_i, denoted by $\widehat{\tau}_{\mathrm{F}}^{\mathrm{cc}}$ and $\widehat{\tau}_{\mathrm{L}}^{\mathrm{cc}}$, respectively, to estimate τ.

In Example 1 with $J = 1$ covariate, we have $C_i = 1 - M_i$, and the complete-case analysis uses only the units with $M_i = 0$. In Example 2 with $J = 2$ covariates, we have $C_i = (1 - M_{i1})(1 - M_{i2})$, and the complete-case analysis uses only the units with $M_{i1} = M_{i2} = 0$.

3.2. Complete-Covariate Analysis

The *complete-covariate analysis* omits any covariates that are not completely observed for all units. Denote by $\mathcal{J} = \{j : M_{ij} = 0 \text{ for all } i = 1, \ldots, N\}$ the set of complete covariates, and let $x_i^{\mathrm{ccov}} = (x_{ij})_{j \in \mathcal{J}}$ be the subvector of x_i corresponding to the covariates in $\mathcal{J}$. The complete-covariate analysis fits

$$Y_i \sim 1 + Z_i + x_i^{\mathrm{ccov}}, \tag{3}$$
$$Y_i \sim 1 + Z_i + (x_i^{\mathrm{ccov}} - \bar{x}^{\mathrm{ccov}}) + Z_i(x_i^{\mathrm{ccov}} - \bar{x}^{\mathrm{ccov}}) \tag{4}$$

over $i = 1, \ldots, N$ under the additive and fully interacted specifications, respectively, and uses the coefficients of Z_i, denoted by $\widehat{\tau}_{\mathrm{F}}^{\mathrm{ccov}}$ and $\widehat{\tau}_{\mathrm{L}}^{\mathrm{ccov}}$, respectively, to estimate τ. In the case of $\mathcal{J} = \emptyset$, both (3) and (4) reduce to $Y_i \sim 1 + Z_i$ with $\widehat{\tau}_{\mathrm{F}}^{\mathrm{ccov}} = \widehat{\tau}_{\mathrm{L}}^{\mathrm{ccov}} = \widehat{\tau}_{\mathrm{N}}$.

3.3. Single Imputation

Single imputation fills in the missing covariates based on the observed data, and analyzes the resulting completed dataset by standard methods. It includes all cases and covariates in the analysis, making full use of the observed data.

Consider a covariate-wise imputation that imputes all missing x_{ij}'s along the jth dimension by some c_j that may depend on the observed data. Common choices include $c_j = 0$ and the covariate-wise observed average $c_j = \widehat{x}_j^{\mathrm{obs}} = \sum_{i=1}^{N}(1 - M_{ij})x_{ij} / \sum_{i=1}^{N}(1 - M_{ij})$. Denote by

$$x_i^{\mathrm{imp}}(c) = (x_{i1}^{\mathrm{imp}}(c_1), \ldots, x_{iJ}^{\mathrm{imp}}(c_J))^{\mathrm{T}}$$

the resulting imputed covariate vector with $c = (c_j)_{j=1}^{J}$ and $x_{ij}^{\mathrm{imp}}(c_j) = (1 - M_{ij})x_{ij} + M_{ij}c_j$. We can proceed with fitting

$$Y_i \sim 1 + Z_i + x_i^{\mathrm{imp}}(c), \tag{5}$$
$$Y_i \sim 1 + Z_i + \{x_i^{\mathrm{imp}}(c) - \bar{x}^{\mathrm{imp}}(c)\} + Z_i\{x_i^{\mathrm{imp}}(c) - \bar{x}^{\mathrm{imp}}(c)\} \tag{6}$$

over $i = 1, \ldots, N$, respectively, and estimate τ by the coefficients of Z_i, denoted by $\widehat{\tau}_{\mathrm{F}}^{\mathrm{imp}}(c)$ and $\widehat{\tau}_{\mathrm{L}}^{\mathrm{imp}}(c)$, respectively.

The covariate-wise imputation fills in identical value for all missing values along the same covariate, and can appear quite restrictive. Other common choices for single imputation

include the treatment-specific sample means of the observed covariates (Schemper and Smith 1990) and other conditional sample means of the observed covariates based on either only the observed covariates or both the observed covariates and outcomes (Little 1992). We will nevertheless focus on the simple covariate-wise imputation given its sufficiency for randomized experiments; see Sullivan et al. (2018), Kayembe et al. (2020), and Kamat and Reiter (2021) for numerical evidence on the insensitivity of standard analyses to imputation methods in randomized experiments. In addition, imputing with treatment-specific sample means is asymptotically equivalent to imputing with $c_j = \widehat{x}_j^{\text{obs}}$ under Condition 1. We leave the theory of more complicated single imputation methods to future work. A key caveat is that imputation based on outcomes may be subject to bias due to adjustment for variables that have been affected by the treatment.

Moreover, we will show that the robust standard errors from OLS are convenient approximations to the true standard errors of the $\widehat{\tau}_*^{\text{imp}}(c)$'s. It is thus unnecessary to resort to multiple imputation, which is a tool to assess uncertainty in general missing data problems (Rubin 1987). We omit it from the following discussion.

3.4. Missingness-Indicator Method

The *missingness-indicator method* augments (5) and (6) under single imputation by also including M_i as additional regressors to account for the missingness information. Specifically, we first impute the missing x_{ij}'s by the covariate-specific c_j's, and then fit

$$Y_i \sim 1 + Z_i + x_i^{\text{imp}}(c) + M_i, \tag{7}$$

$$Y_i \sim 1 + Z_i + \{x_i^{\text{imp}}(c) - \bar{x}^{\text{imp}}(c)\} + (M_i - \bar{M}) + Z_i\{x_i^{\text{imp}}(c) - \bar{x}^{\text{imp}}(c)\} + Z_i(M_i - \bar{M}) \tag{8}$$

over $i = 1, \ldots, N$ to construct the regression estimators as the coefficients of Z_i, denoted by $\widehat{\tau}_{\text{F}}^{\text{mim}}(c)$ and $\widehat{\tau}_{\text{L}}^{\text{mim}}(c)$, respectively. This is equivalent to fitting the additive and fully interacted regressions based on the augmented covariate vector $x_i^{\text{mim}}(c) = (x_i^{\text{imp}}(c)^{\text{T}}, M_i^{\text{T}})^{\text{T}}$.

Recall that $\mathcal{J}$ denotes the set of complete covariates. Strictly speaking, $M_{ij} = 0$ for all $i = 1, \ldots, N$ for $j \in \mathcal{J}$ such that we need to include only the subvector of M_i that corresponds to the incomplete covariates. In addition, in case $M_{ij} = M_{ij'}$ for all $i = 1, \ldots, N$ for some $j \neq j'$ outside $\mathcal{J}$, we need to include only one of them to avoid collinearity. Acknowledging the need to adjust for these complications on a case-by-case basis, we will use M_i to represent the vector of missingness indicators added to the regressions after appropriate adjustment for notational simplicity. This causes little confusion because standard software packages for OLS automatically drop redundant regressors.

As it turns out, $\widehat{\tau}_*^{\text{mim}}(c)$ $(* = \text{F}, \text{L})$ and their associated robust standard errors $\widehat{\text{se}}_*^{\text{mim}}(c)$ $(* = \text{F}, \text{L})$ are all invariant to the choice of the imputation vector c. This is a numeric merit due to the inclusion of the missingness indicators, and allows us to construct $\widehat{\tau}_*^{\text{mim}}(c)$ by simply imputing all missing covariates as 0. We formalize the intuition in Theorem 1 below.

Theorem 1. $\widehat{\tau}_*^{\text{mim}}(c)$ and $\widehat{\text{se}}_*^{\text{mim}}(c)$ are invariant to the choice of $c \in \mathbb{R}^J$ for $* = \text{F}, \text{L}$.

Cohen and Cohen (1975) hinted at the invariance of $\widehat{\tau}_{\text{F}}^{\text{mim}}(c)$ to the choice of c. Theorem 1 formalizes the results for both $\widehat{\tau}_{\text{F}}^{\text{mim}}(c)$ and $\widehat{\tau}_{\text{L}}^{\text{mim}}(c)$, as well as the corresponding robust standard errors. We will use $\widehat{\tau}_*^{\text{mim}}$ and $\widehat{\text{se}}_*^{\text{mim}}$ to denote the common values of $\widehat{\tau}_*^{\text{mim}}(c)$ and $\widehat{\text{se}}_*^{\text{mim}}(c)$ across all c. As it turns out, augmenting $x_i^{\text{imp}}(c)$ by M_i is not only sufficient but also necessary to ensure the invariance of the resulting regression estimators to the imputed values. We formalize the result in Theorem S1 of the supplementary materials.

3.5. Missingness-Pattern Method

The missingness-indicator method factors in information in missingness by including M_i as additional regressors in OLS. We propose the *missingness-pattern method* that goes one step further, and fits one separate regression for each missingness pattern based on all available covariates. Let $\rho_{(m)} = N_{(m)}/N$ denote the proportion of units with missingness pattern $m \in \{0, 1\}^J$. We use Examples 1 and 2 (continued) below to illustrate the basic idea with $J = 1$ and $J = 2$, respectively, and then formalize the method for general J.

Example 1 (continued). Consider the case of $J = 1$ covariate, $x_i \in \mathbb{R}$, and two missingness patterns, $m \in \mathcal{M} = \{0, 1\}$. We can fit one additive regression for each missingness pattern to obtain the coefficients of Z_i as follows:

1. regress Y_i on $(1, Z_i, x_i)$ over $\{i : M_i = 0\}$ with x_i observed to obtain $\widehat{\tau}_{\text{F},(0)}$;
2. regress Y_i on $(1, Z_i)$ over $\{i : M_i = 1\}$ with x_i missing to obtain $\widehat{\tau}_{\text{F},(1)} = \widehat{\tau}_{\text{N},(1)}$.

The weighted average $\widehat{\tau}_{\text{F}}^{\text{mp}} = \rho_{(0)}\widehat{\tau}_{\text{F},(0)} + \rho_{(1)}\widehat{\tau}_{\text{F},(1)}$ gives an estimator for τ. Analogously, we can construct $\widehat{\tau}_{\text{L}}^{\text{mp}}$ by fitting one fully interacted regression for each missingness pattern.

Example 2 (continued). Consider the case of $J = 2$ covariates, $x_i = (x_{i1}, x_{i2})^{\text{T}}$, and four missingness patterns, $m \in \mathcal{M} = \{0, 1\}^2$. We can fit one additive regression for each missingness pattern to obtain the coefficients of Z_i as follows:

1. regress Y_i on $(1, Z_i, x_{i1}, x_{i2})$ over $\{i : M_i = (0, 0)^{\text{T}}\}$ with both covariates observed to obtain $\widehat{\tau}_{\text{F},(0,0)}$;
2. regress Y_i on $(1, Z_i, x_{i1})$ over $\{i : M_i = (0, 1)^{\text{T}}\}$ with only covariate 1 observed to obtain $\widehat{\tau}_{\text{F},(0,1)}$;
3. regress Y_i on $(1, Z_i, x_{i2})$ over $\{i : M_i = (1, 0)^{\text{T}}\}$ with only covariate 2 observed to obtain $\widehat{\tau}_{\text{F},(1,0)}$;
4. regress Y_i on $(1, Z_i)$ over $\{i : M_i = (1, 1)^{\text{T}}\}$ with both covariates missing to obtain $\widehat{\tau}_{\text{F},(1,1)} = \widehat{\tau}_{\text{N},(1,1)}$.

The weighted average $\widehat{\tau}_{\text{F}}^{\text{mp}} = \rho_{(0,0)}\widehat{\tau}_{\text{F},(0,0)} + \rho_{(0,1)}\widehat{\tau}_{\text{F},(0,1)} + \rho_{(1,0)}\widehat{\tau}_{\text{F},(1,0)} + \rho_{(1,1)}\widehat{\tau}_{\text{F},(1,1)}$ gives an estimator of τ. Analogously, we can construct $\widehat{\tau}_{\text{L}}^{\text{mp}}$ by fitting one fully interacted regression for each missingness pattern.

Extensions to general J are immediate. Let $x_i^{\text{mp}} = (x_{ij})_{j:M_{ij}=0}$ be the vector of available covariates for unit i. In the above Example 2 (continued), we have (i) $x_i^{\text{mp}} = x_i = (x_{i1}, x_{i2})^{\text{T}}$ for units with $M_i = (0, 0)^{\text{T}}$; (ii) $x_i^{\text{mp}} = x_{i1}$ for units with $M_i = (0, 1)^{\text{T}}$; (iii) $x_i^{\text{mp}} = x_{i2}$ for units with $M_i = (1, 0)^{\text{T}}$; and (iv) $x_i^{\text{mp}} = \emptyset$ for units with $M_i = (1, 1)^{\text{T}}$. For units with

missingness pattern $m \in \mathcal{M}$, the missingness-pattern method fits

$$Y_i \sim 1 + Z_i + x_i^{\text{mp}}, \tag{9}$$

$$Y_i \sim 1 + Z_i + (x_i^{\text{mp}} - \bar{x}_{(m)}^{\text{mp}}) + Z_i(x_i^{\text{mp}} - \bar{x}_{(m)}^{\text{mp}}) \tag{10}$$

over $\{i : M_i = m\}$ under the additive and fully interacted specifications, respectively, where $\bar{x}_{(m)}^{\text{mp}} = N_{(m)}^{-1} \sum_{i:M_i=m} x_i^{\text{mp}}$. Let $\widehat{\tau}_{\text{F},(m)}$ and $\widehat{\tau}_{\text{L},(m)}$ be the coefficients of Z_i from (9) and (10), respectively. The weighted averages

$$\widehat{\tau}_*^{\text{mp}} = \sum_{m \in \mathcal{M}} \rho_{(m)} \widehat{\tau}_{*,(m)} \qquad (* = \text{F}, \text{L}) \tag{11}$$

give two covariate-adjusted estimators of τ.

The missingness-pattern method differentiates between all $|\mathcal{M}|$ missingness patterns like the missingness-indicator method, yet does so by fitting $|\mathcal{M}|$ missingness-pattern-specific OLS. It can be seen as a hybrid of the complete-case and complete-covariate analyses, factoring in all available covariates without the need of imputation or augmentation. With $\widehat{\tau}_{*,(0_J)}$ coinciding with $\widehat{\tau}_*^{\text{cc}}$ from the complete-case analysis, it can also be seen as an ensemble variant of $\widehat{\tau}_*^{\text{cc}}$, averaging over estimators based on not only the complete cases but also other missingness patterns present in the dataset.

The idea of missingness-pattern-specific analysis dates back to Wilks (1932), Matthai (1951), and Rosenbaum and Rubin (1984, Appendix B), and is reminiscent of the pattern-mixture model approach to handling missingness that formulates distinct models within each missingness pattern (Little 1993). Yet its use for analyzing experiments with missing covariates remains mostly unexploited to the best of our knowledge. A key intuition is that the missingness pattern acts as a discrete pretreatment covariate, and thus allows for post-stratified estimators by averaging over estimators within missingness patterns. Miratrix, Sekhon, and Yu (2013) demonstrated the asymptotic efficiency gain of post-stratification based on the simple stratum-specific differences in means without adjusting for additional covariates. The $\widehat{\tau}_{\text{L}}^{\text{mp}}$ in (11) averages over regression-adjusted estimators within missingness patterns, and promises additional large-sample efficiency over the missingness indicator method by allowing heterogeneous adjustments across different missingness patterns. We quantify the intuition in Section 4. Importantly, unlike the pattern-mixture model approach that models the joint distribution of the missing covariates with the observed covariates and outcome within each missingness pattern, the missingness-pattern method in (9) and (10) bases inference on the available covariates within each stratum, and makes no assumption on the distribution of either the covariates or outcome.

3.6. Summary of the Regression-Adjusted Estimators

Sections 3.1–3.5 present in total ten regression-adjusted estimators, $\widehat{\tau}_*^\dagger$, as the combinations of five missing-data strategies, $\dagger \in \{\text{cc}, \text{ccov}, \text{imp}, \text{mim}, \text{mp}\}$, and two model specifications, $* = \text{F}, \text{L}$. Table 1 summarizes them. Of interest is their respective validity and efficiency for inferring τ. A key observation is that $\emptyset \subseteq x_i^{\text{ccov}} \subseteq x_i^{\text{imp}}(c) \subseteq x_i^{\text{mim}}(c)$ for arbitrary $c \in \mathbb{R}^J$. By Remark 1, this elucidates the asymptotic efficiency of $\widehat{\tau}_{\text{L}}^{\text{mim}}$ over $\widehat{\tau}_{\text{L}}^{\text{imp}}(c)$, $\widehat{\tau}_{\text{L}}^{\text{ccov}}$, and $\widehat{\tau}_{\text{N}}$ if the $x_i^\dagger$'s act as standard covariates. We formalize the intuition in Section 4.

4. Design-based Theory

We quantify in this section the design-based properties of the estimators in Table 1. In particular, regression adjustment delivers not only point estimators but also their associated robust standard errors, denoted by $\widehat{\text{se}}_*^\dagger$. We focus on the validity of $(\widehat{\tau}_*^\dagger, \widehat{\text{se}}_*^\dagger)$ for large-sample Wald-type inference, which concerns the point and interval estimation of τ based on not only the consistency and asymptotic normality of $\widehat{\tau}_*^\dagger$ but also the asymptotic conservativeness of $\widehat{\text{se}}_*^\dagger$ for estimating the true standard error. In brief, we do not recommend $\widehat{\tau}_*^{\text{cc}}$ $(* = \text{F}, \text{L})$ due to their inconsistency without a strong additional assumption. We recommend $\widehat{\tau}_{\text{L}}^{\text{mim}}$ in general due to its simplicity, invariance to c, and asymptotic efficiency over $\{\widehat{\tau}_{\text{N}}, \widehat{\tau}_{\text{F}}^{\text{mim}}, \widehat{\tau}_*^{\text{ccov}}, \widehat{\tau}_*^{\text{imp}}(c) : * = \text{F}, \text{L}\}$. When the missingness-pattern-specific sample sizes permit, we recommend $\widehat{\tau}_{\text{L}}^{\text{mp}}$ due to its additional gain in asymptotic efficiency. Echoing Remark 1, a key caveat is that the asymptotic efficiency of $\widehat{\tau}_{\text{L}}^{\text{mim}}$ and $\widehat{\tau}_{\text{L}}^{\text{mp}}$ may come at the cost of finite sample performance. Importantly, despite $\widehat{\tau}_*^\dagger$ and $\widehat{\text{se}}_*^\dagger$ are originally motivated by regression models, our theory is design-based and holds even when the regression models are misspecified.

With a slight redundancy of notation, let $A_{ij} = 1 - M_{ij}$ indicate the availability of x_{ij} with $A_i = (A_{i1}, \ldots, A_{iJ})^{\text{T}} = 1_J - M_i$ and $A_i \circ x_i = (A_{i1}x_{i1}, \ldots, A_{iJ}x_{iJ})^{\text{T}} = x_i^{\text{imp}}(0_J)$. We need the following regularity conditions for asymptotics under

Table 1. Ten regression-adjusted estimators $\widehat{\tau}_*^\dagger$ with $\dagger \in \{\text{cc}, \text{ccov}, \text{imp}, \text{mim}, \text{mp}\}$ and $* = \text{F}, \text{L}$.

$\widehat{\tau}_*^\dagger$	Missing-covariate strategy	Covariates for regressions	Consistency
$\widehat{\tau}_*^{\text{cc}}$	use complete cases and all covariates	x_i	No
$\widehat{\tau}_*^{\text{ccov}}$	use all units and complete covariates	$x_i^{\text{ccov}} = (x_{ij})_{j \in \mathcal{J}}$	Yes
$\widehat{\tau}_*^{\text{imp}}(c)$	impute the missing x_{ij}'s with c_j; run ols with all covariates	$x_i^{\text{imp}}(c) = (x_{i1}^{\text{imp}}(c_1), \ldots, x_{iJ}^{\text{imp}}(c_J))^{\text{T}}$ with $x_{ij}^{\text{imp}}(c_j) = (1 - M_{ij})x_{ij} + M_{ij}c_j$	Yes
$\widehat{\tau}_*^{\text{mim}}$	impute the missing x_{ij}'s with 0; augment the covariates with M_{ij}'s	$x_i^{\text{mim}}(0_J) = (x_i^{\text{imp}}(0_J)^{\text{T}}, M_i^{\text{T}})^{\text{T}}$	Yes
$\widehat{\tau}_*^{\text{mp}}$	run missingness-pattern-specific OLS	$x_i^{\text{mp}} = (x_{ij})_{j:M_{ij}=0}$	Yes

NOTE: The complete-case analysis uses units with $C_i = 1$, whereas the other four strategies use all units. Under suitable regularity conditions, $\widehat{\tau}_*^{\text{cc}}$ $(* = \text{F}, \text{L})$ can be inconsistent, whereas the other eight estimators are consistent. Among the eight consistent estimators, (i) $\widehat{\tau}_{\text{L}}^\dagger$ is asymptotically more efficient than $\widehat{\tau}_{\text{F}}^\dagger$ and $\widehat{\tau}_{\text{N}}$ for all $\dagger \in \{\text{ccov}, \text{imp}, \text{mim}, \text{mp}\}$; (ii) $\widehat{\tau}_{\text{L}}^{\text{mim}}$ is asymptotically the most efficient among $\{\widehat{\tau}_{\text{N}}, \widehat{\tau}_*^{\text{ccov}}, \widehat{\tau}_*^{\text{imp}}(c), \widehat{\tau}_*^{\text{mim}} : * = \text{F}, \text{L}\}$; (iii) $\widehat{\tau}_{\text{L}}^{\text{mp}}$ is asymptotically more efficient than $\widehat{\tau}_{\text{L}}^{\text{mim}}$, but can have large variability in finite samples.

the complete-case, complete-covariate, single imputation, and missingness-indicator strategies.

Condition 2. Assume Condition 1. As $N \to \infty$, (i) e_z has a limit in $(0,1)$ for $z = 0, 1$, (ii) $\mathcal{J}$ has a limit, (iii) the first two finite population moments of $\{Y_i(z), x_i, M_i, C_i, C_iY_i(z), C_ix_i, A_i \circ x_i : z = 0, 1\}_{i=1}^N$ have finite limits; the limit of $\bar{C} = N^{-1}\sum_{i=1}^N C_i$ is in $(0, 1]$, and (iv) $N^{-1}\sum_{i=1}^N \|x_i\|_4^4 = O(1)$ and $N^{-1}\sum_{i=1}^N Y_i^4(z) = O(1)$ for $z = 0, 1$.

Condition 2 gives the union of the regularity conditions required by individual strategies to facilitate comparison and avoid repetition. The conditions on C_i are required by only the complete-case analysis. The condition on $\mathcal{J}$ is required by only the complete-covariate analysis. The conditions on M_i and $A_i \circ x_i$ are required by only the single imputation and missingness-indicator method.

4.1. Complete-Case Analysis

We derive in this section the design-based properties of $(\widehat{\tau}_*^{\text{cc}}, \widehat{\text{se}}_*^{\text{cc}})$ $(* = \text{F}, \text{L})$ from (1) and (2). In brief, $\widehat{\tau}_*^{\text{cc}}$ is in general not consistent for τ unless the average treatment effect of the complete cases equals that of the incomplete cases asymptotically. This is a strong assumption, and can be problematic whenever missingness is correlated with the potential outcomes. A common example in randomized clinical trials is that the more severely ill patients are often more likely to have missing pretreatment covariates. We therefore do not recommend the complete-case analysis. We present the technical details below.

Recall that $\{i : C_i = 1\}$ denotes the set of the N^{cc} complete cases. Let $\tau^{\text{cc}} = (N^{\text{cc}})^{-1}\sum_{i:C_i=1}\tau_i$ denote the corresponding average treatment effect, which is nonstochastic and has a finite limit under Condition 2. Let S_{xx}^{cc} be the finite population covariance of $(x_i)_{i:C_i=1}$, which has a finite limit under Condition 2. Let v_*^{cc} and $S_{\tau\tau,*}^{\text{cc}}$ be the analogs of v_* and $S_{\tau,*}^2$ in Lemma 1 defined over $\{Y_i(0), Y_i(1), x_i\}_{i:C_i=1}$ for $* = \text{F}, \text{L}$.

Proposition 1. Assume complete randomization, Condition 2, and the limit of S_{xx}^{cc} is nonsingular. Then $\sqrt{N^{\text{cc}}}(\widehat{\tau}_*^{\text{cc}} - \tau^{\text{cc}}) \rightsquigarrow \mathcal{N}(0, v_*^{\text{cc}})$ with $v_{\text{L}}^{\text{cc}} \le v_{\text{F}}^{\text{cc}}$. In addition, $N^{\text{cc}}(\widehat{\text{se}}_*^{\text{cc}})^2 - v_*^{\text{cc}} = S_{\tau\tau,*}^{\text{cc}} + o_{\text{P}}(1)$, where $S_{\tau\tau,*}^{\text{cc}} \ge 0$.

Proposition 1 justifies the use of $(\widehat{\tau}_*^{\text{cc}}, \widehat{\text{se}}_*^{\text{cc}})$ for the large-sample Wald-type inference of τ^{cc}. The complete-case analysis is consistent for τ if and only if $\tau^{\text{cc}} - \tau = o(1)$. This imposes a strong restriction on the missingness mechanism, which in general does not hold.

4.2. Complete-Covariate Analysis

Recall that $x_i^{\text{ccov}} = (x_{ij})_{j\in\mathcal{J}}$ denotes the vector of complete covariates used by the complete-covariate analysis. The finite population covariance of $(x_i^{\text{ccov}})_{i=1}^N$, denoted by S_{xx}^{ccov}, has a finite limit under Condition 2. Let v_*^{ccov} and $S_{\tau\tau,*}^{\text{ccov}}$ be the analogs of v_* and $S_{\tau,*}^2$ in Lemma 1 defined over $\{Y_i(0), Y_i(1), x_i^{\text{ccov}}\}_{i=1}^N$ for $* = \text{F}, \text{L}$.

Proposition 2. Assume complete randomization, Condition 2, and the limit of S_{xx}^{ccov} is nonsingular. Then $\sqrt{N}(\widehat{\tau}_*^{\text{ccov}} - \tau) \rightsquigarrow \mathcal{N}(0, v_*^{\text{ccov}})$ with $v_{\text{L}}^{\text{ccov}} \le v_{\text{F}}^{\text{ccov}}$ and $v_{\text{L}}^{\text{ccov}} \le v_{\text{N}}$. In addition, $N(\widehat{\text{se}}_*^{\text{ccov}})^2 - v_*^{\text{ccov}} = S_{\tau\tau,*}^{\text{ccov}} + o_{\text{P}}(1)$, where $S_{\tau\tau,*}^{\text{ccov}} \ge 0$.

Proposition 2 justifies the large-sample Wald-type inference based on $(\widehat{\tau}_*^{\text{ccov}}, \widehat{\text{se}}_*^{\text{ccov}})$ regardless of whether $\tau^{\text{cc}} = \tau$ or not, and ensures the asymptotic efficiency of $\widehat{\tau}_{\text{L}}^{\text{ccov}}$ over $\widehat{\tau}_{\text{F}}^{\text{ccov}}$ and the unadjusted $\widehat{\tau}_{\text{N}}$. This illustrates the advantage of including all units in the analysis even at the cost of discarding all information in the incomplete covariates. Importantly, all theoretical guarantees hold even when the missingness is related to the missing covariates and unobserved potential outcomes, a scenario analogous to missing not at random under the superpopulation framework (Rubin 1976).

Schemper and Smith (1990) referred to the complete-covariate analysis as "an even less justifiable method" than the complete-case analysis. Whereas their comment could be valid for observational studies when incomplete covariates include some key confounders, we give the opposite recommendation for randomized experiments. Intuitively, randomization precludes the possibility of confounding by enforcing independence between the treatment assignment and pretreatment covariates, ensuring valid simple comparisons even without covariate adjustment. The exclusion of incomplete covariates thus does not affect the validity of the complete-covariate analysis while allowing for additional asymptotic efficiency over $\widehat{\tau}_{\text{N}}$ as long as one complete covariate is prognostic. We consider this as the baseline strategy for benchmarking the alternative strategies.

4.3. Single Imputation

Under single imputation, we allow the imputation vector $c = (c_j)_{j=1}^J$ to be dependent on $(Z_i)_{i=1}^N$, and hence stochastic from the design-based perspective. An example is $c_j = \sum_{i:Z_i=1}(1 - M_{ij})x_{ij} / \sum_{i:Z_i=1}(1 - M_{ij})$, as the sample means of the observed covariates under treatment level 1. We focus on c's that have finite probability limits under complete randomization:

$$\mathcal{C} = \{c \in \mathbb{R}^J : \text{plim } c = c_\infty < \infty \text{ under complete randomization and Condition 2}\}.$$

The constant imputation with $c_j = 0$ for $j = 1, \ldots, J$ is a special case with $c_\infty = c = 0_J$. The unconditional sample mean imputation with $c_j = \bar{x}_j^{\text{obs}}$ is a special case with $c_\infty = (c_{j,\infty})_{j=1}^J$, where $c_{j,\infty} = \text{plim } c_j = \lim_{N\to\infty}\sum_{i=1}^N A_{ij}x_{ij} / \sum_{i=1}^N A_{ij}$.

Let $S_{xx}^{\text{imp}}(c)$ be the finite population covariance of $\{x_i^{\text{imp}}(c)\}_{i=1}^N$. Condition 2 ensures that $S_{xx}^{\text{imp}}(c)$ has a finite limit for all $c \in \mathcal{C}$. Let $v_*^{\text{imp}}(c_\infty)$ and $S_{\tau\tau,*}^{\text{imp}}(c_\infty)$ be the analogs of v_* and $S_{\tau,*}^2$ in Lemma 1 defined over $\{Y_i(0), Y_i(1), x_i^{\text{imp}}(c_\infty)\}_{i=1}^N$ for $* = \text{F}, \text{L}$.

Proposition 3. Assume complete randomization, Condition 2, and $c \in \mathcal{C}$ with the limit of $S_{xx}^{\text{imp}}(c)$ being nonsingular. Then $\sqrt{N}\{\widehat{\tau}_*^{\text{imp}}(c) - \tau\} \rightsquigarrow \mathcal{N}\{0, v_*^{\text{imp}}(c_\infty)\}$ with $v_{\text{L}}^{\text{imp}}(c_\infty) \le v_{\text{F}}^{\text{imp}}(c_\infty)$ and $v_{\text{L}}^{\text{imp}}(c_\infty) \le v_{\text{L}}^{\text{ccov}} \le v_{\text{N}}$. In addition, $N\{\widehat{\text{se}}_*^{\text{imp}}(c)\}^2 - v_*^{\text{imp}}(c_\infty) = S_{\tau\tau,*}^{\text{imp}}(c_\infty) + o_{\text{P}}(1)$, where $S_{\tau\tau,*}^{\text{imp}}(c_\infty) \ge 0$.

Echoing the comments after Proposition 2, Proposition 3 justifies the large-sample Wald-type inference based on $\{\widehat{\tau}_*^{\text{imp}}(c), \widehat{\text{se}}_*^{\text{imp}}(c)\}$ for $* = \text{F}, \text{L}$ without any restrictions on the relation between M_i's and $\{Y_i(0), Y_i(1), x_i\}_{i=1}^N$ beyond Condition 2. Let $x_i^{\text{imp}}(c_\infty)$ denote the counterpart of $x_i^{\text{imp}}(c)$ if we impute by the finite probability limit of c. The regularity conditions in Proposition 3 guarantee that the $x_i^{\text{imp}}(c_\infty)$'s act as standard covariates, and ensure that $\widehat{\tau}_*^{\text{imp}}(c_\infty)$ is asymptotically normal with mean τ and variance $v_*^{\text{imp}}(c_\infty)$ by Lemma 1, with $v_{\text{L}}^{\text{imp}}(c_\infty) \leq v_{\text{F}}^{\text{imp}}(c_\infty)$. The proof of Proposition 3 in the supplementary materials further ensures that $\widehat{\tau}_*^{\text{imp}}(c)$ has the same limiting distribution as that of $\widehat{\tau}_*^{\text{imp}}(c_\infty)$. This explains the somehow surprising result that the possible randomness in c, if any, does not increase the asymptotic variances of $\widehat{\tau}_*^{\text{imp}}(c)$'s, and ensures the consistency of $\widehat{\tau}_*^{\text{imp}}(c)$'s along with the efficiency of $\widehat{\tau}_{\text{L}}^{\text{imp}}(c)$ over $\widehat{\tau}_{\text{F}}^{\text{imp}}(c)$. The efficiency of $\widehat{\tau}_{\text{L}}^{\text{imp}}(c)$ over $\widehat{\tau}_*^{\text{ccov}}$ $(* = \text{F}, \text{L})$ and $\widehat{\tau}_{\text{N}}$ then follows from Remark 1 with $x_i^{\text{imp}}(c) \supseteq x_i^{\text{ccov}} \supseteq \emptyset$. This suggests the advantage of including all covariates in the analysis even with some basic imputation.

Observe that the true and estimated variances both depend on the choice of c. We can minimize them over c to obtain the optimal imputation. However, we do not go into details of this route because the imputation method is strictly dominated by the missingness-indicator method, as shown in the next section.

4.4. Missingness-Indicator Method

Recall that $x_i^{\text{mim}}(0_J) = (x_i^{\text{imp}}(0_J)^{\text{T}}, M_i^{\text{T}})^{\text{T}}$ denotes the covariates for forming (7) and (8) under the missingness-indicator method with all missing covariates imputed by 0. Let S_{xx}^{mim} be the finite population covariance of $\{x_i^{\text{mim}}(0_J)\}_{i=1}^N$. It has a finite limit under Condition 2. Let v_*^{mim} and $S_{\tau\tau,*}^{\text{mim}}$ be the analogs of v_* and $S_{\tau,*}^2$ in Lemma 1 defined over $\{Y_i(0), Y_i(1), x_i^{\text{mim}}(0_J)\}_{i=1}^N$ for $* = \text{F}, \text{L}$. Recall that $\widehat{\tau}_*^{\text{mim}}$ and $\widehat{\text{se}}_*^{\text{mim}}$ denote the common values of $\widehat{\tau}_*^{\text{mim}}(c)$ and $\widehat{\text{se}}_*^{\text{mim}}(c)$ across all $c \in \mathbb{R}^J$ by Theorem 1.

Proposition 4. Assume complete randomization, Condition 2, and the limit of S_{xx}^{mim} is nonsingular. Then $\sqrt{N}(\widehat{\tau}_*^{\text{mim}} - \tau) \rightsquigarrow \mathcal{N}(0, v_*^{\text{mim}})$ with $v_{\text{L}}^{\text{mim}} \leq v_{\text{F}}^{\text{mim}}$ and $v_{\text{L}}^{\text{mim}} \leq v_{\text{L}}^{\text{imp}}(c_\infty) \leq v_{\text{L}}^{\text{ccov}} \leq v_{\text{N}}$ for all $c \in \mathcal{C}$. In addition, $N(\widehat{\text{se}}_*^{\text{mim}})^2 - v_*^{\text{mim}} = S_{\tau\tau,*}^{\text{mim}} + o_{\text{P}}(1)$, where $S_{\tau\tau,*}^{\text{mim}} \geq 0$.

Similar to Propositions 2 and 3, Proposition 4 ensures the validity of the large-sample Wald-type inference based on $(\widehat{\tau}_*^{\text{mim}}, \widehat{\text{se}}_*^{\text{mim}})$ $(* = \text{F}, \text{L})$ without any restrictions on the relation between M_i's and $\{Y_i(0), Y_i(1), x_i\}_{i=1}^N$ beyond Condition 2. Intuitively, randomization balances both covariates and missingness indicators across treatment groups, and thereby ensures consistency of the regression adjustments based on them regardless of the missingness mechanism. The asymptotic efficiency of $\widehat{\tau}_{\text{L}}^{\text{mim}}$ over $\widehat{\tau}_{\text{F}}^{\text{mim}}$, $\widehat{\tau}_{\text{L}}^{\text{imp}}(c)$, $\widehat{\tau}_{\text{L}}^{\text{ccov}}$, and $\widehat{\tau}_{\text{N}}$ then follows from Lemma 1 and Remark 1 with $x_i^{\text{mim}}(c) \supseteq x_i^{\text{imp}}(c) \supseteq x_i^{\text{ccov}} \supseteq \emptyset$. This highlights a second advantage of including the missingness indicators in addition to the invariance to imputed values. Importantly, observe that the missingness indicator method essentially doubles the number of covariates relative to single imputation, whereas the fully interacted specification doubles model complexity relative to the additive specification. The asymptotic efficiency of $\widehat{\tau}_{\text{L}}^{\text{mim}}$ may thus come at the price of high variance in finite samples, echoing Remark 1. We illustrate this point using simulation in the supplementary materials.

Remark 3. Jones (1996) assumed that the outcome follows the Gauss–Markov model $Y_i = \mu + Z_i\beta + x_i\gamma + \epsilon_i$, with x_i being a univariate covariate that is possibly missing and β being the constant treatment effect that gives the model-based analog of τ. He showed that $\widehat{\tau}_{\text{F}}^{\text{mim}}$ is unbiased for β if the sample covariance of Z_i and x_i for those units with missing x_i equals zero. Complete randomization ensures that this sample covariance converges to zero in probability, and guarantees the consistency of $\widehat{\tau}_{\text{F}}^{\text{mim}}$. Importantly, Jones's (1996) theory is model-based, whereas ours is design-based.

4.5. Missingness-Pattern Method

4.5.1. Conditional Properties under Post-Stratification

Recall that $\mathcal{M} = \{m : N_{(m)} > 0\}$ denotes the set of missingness patterns present in the study population, with $\rho_{(m)} = N_{(m)}/N$ as the proportion of units with missingness pattern m. Let $\widehat{\text{se}}_{*,(m)}$ be the robust standard error associated with $\widehat{\tau}_{*,(m)}$ from the pattern-specific OLS fit under missingness pattern $m \in \mathcal{M}$. The weighted average

$$(\widehat{\text{se}}_*^{\text{mp}})^2 = \sum_{m \in \mathcal{M}} \rho_{(m)}^2 \widehat{\text{se}}_{*,(m)}^2 \tag{12}$$

gives an intuitive estimator of the sampling variance of $\widehat{\tau}_*^{\text{mp}}$ from (11) for $* = \text{F}, \text{L}$.

Denote by $N_{(m),z}$ the number of units with missingness pattern m that receive treatment level $z \in \{0, 1\}$. Conditioning on $\mathcal{D} = \{N_{(m),z} : m \in \mathcal{M}, z = 0, 1\}$ with $N_{(m),z} > 0$ for all $m \in \mathcal{M}$ and $z = 0, 1$, we have $|\mathcal{M}|$ independent completely randomized experiments, one within each missingness pattern (Miratrix, Sekhon, and Yu 2013). Under regularity conditions within each missingness pattern, Lemma 1 ensures the asymptotic normality of $\widehat{\tau}_{*,(m)}$ $(* = \text{F}, \text{L})$, the asymptotic efficiency of $\widehat{\tau}_{\text{L},(m)}$ over $\widehat{\tau}_{\text{F},(m)}$, and the asymptotic conservativeness of $\widehat{\text{se}}_{*,(m)}$ for estimating the true standard error of $\widehat{\tau}_{*,(m)}$ for $m \in \mathcal{M}$. Consequently, $\widehat{\tau}_*^{\text{mp}}$ is asymptotically normal for $* = \text{F}, \text{L}$, with $\widehat{\tau}_{\text{L}}^{\text{mp}}$ being asymptotically more efficient than $\widehat{\tau}_{\text{F}}^{\text{mp}}$ and the $\widehat{\text{se}}_*^{\text{mp}}$'s being asymptotically conservative for estimating their respective true standard errors.

The above conditional theory for the missingness-pattern method is straightforward and elegant. A fair comparison with other methods, however, requires quantification of its asymptotic behaviors without conditioning on $\mathcal{D}$. We address this in Section 4.5.2 below.

4.5.2. Unconditional Properties via Aggregate Regression

The unconditional theory would be intuitive if we can express $(\widehat{\tau}_*^{\text{mp}}, \widehat{\text{se}}_*^{\text{mp}})$ as outputs from one aggregate regression. As it turns out, regression adjustment with $x_i^{\text{imp}}(c), M_{i1}, \ldots, M_{iJ}$, and all their interactions recovers $(\widehat{\tau}_{\text{F}}^{\text{mp}}, \widehat{\text{se}}_{\text{F}}^{\text{mp}})$ and $(\widehat{\tau}_{\text{L}}^{\text{mp}}, \widehat{\text{se}}_{\text{L}}^{\text{mp}})$ via one aggregate OLS fit each. We formalize below the intuition for $(\widehat{\tau}_{\text{L}}^{\text{mp}}, \widehat{\text{se}}_{\text{L}}^{\text{mp}})$, and relegate the analogous results for $(\widehat{\tau}_{\text{F}}^{\text{mp}}, \widehat{\text{se}}_{\text{F}}^{\text{mp}})$ to the supplementary materials.

Let $u_i^{\text{mp}}(c)$ be the covariate vector that includes $x_i^{\text{imp}}(c)$, $M_{i1},\ldots,M_{iJ}$, and all their interactions up to some adjustment for collinearity. We give the explicit forms of $u_i^{\text{mp}}(c)$ for $J=1,2$ in Examples 1 and 2 (continued) below, and then state its utility for recovering $(\widehat{\tau}_{\text{L}}^{\text{mp}},\widehat{\text{se}}_{\text{L}}^{\text{mp}})$ via one aggregate regression in Theorem 2.

Example 1 (continued). For $J=1$ with $M_i=M_{i1}$ and $x_i^{\text{imp}}(c)=(1-M_i)x_i+M_ic$, we have

$$u_i^{\text{mp}}(c)=\{x_i^{\text{imp}}(c),M_i,x_i^{\text{imp}}(c)M_i\}=\{x_i^{\text{imp}}(c),M_i\},$$

where the last equality follows from that $x_i^{\text{imp}}(c)M_i=M_ic$ is collinear with M_i. This ensures $u_i^{\text{mp}}(c)=x_i^{\text{mim}}(c)$.

Example 2 (continued). For $J=2$ with $M_i=(M_{i1},M_{i2})^{\text{T}}$ and $x_i^{\text{imp}}(c)=(x_{i1}^{\text{imp}}(c_1),x_{i2}^{\text{imp}}(c_2))^{\text{T}}$, where $x_{ij}^{\text{imp}}(c_j)=(1-M_{ij})x_{ij}+M_{ij}c_j$ $(j=1,2)$, we have

$$\begin{aligned}u_i^{\text{mp}}(c)&=\{x_i^{\text{imp}}(c),M_{i1},M_{i2},M_{i1}M_{i2},x_i^{\text{imp}}(c)M_{i1},\\&\qquad x_i^{\text{imp}}(c)M_{i2},x_i^{\text{imp}}(c)M_{i1}M_{i2}\}\\&=\{x_i^{\text{imp}}(c),M_{i1},M_{i2},M_{i1}M_{i2},x_{i1}^{\text{imp}}(c_1)M_{i2},x_{i2}^{\text{imp}}(c_2)M_{i1}\}.\end{aligned}$$

The last equality follows from that $x_{ij}^{\text{imp}}(c_j)M_{ij}=M_{ij}c_j$ is collinear with M_{ij} for $j=1,2$, and $x_i^{\text{imp}}(c)M_{i1}M_{i2}=(c_1,c_2)^{\text{T}}M_{i1}M_{i2}$ is collinear with $M_{i1}M_{i2}$.

Theorem 2. The fully interacted missingness-pattern estimators $\widehat{\tau}_{\text{L}}^{\text{mp}}$ and $\widehat{\text{se}}_{\text{L}}^{\text{mp}}$ from (11) and (12) equal the coefficient of Z_i and its associated robust standard error from

$$Y_i\sim 1+Z_i+\{u_i^{\text{mp}}(c)-\bar{u}^{\text{mp}}(c)\}+Z_i\{u_i^{\text{mp}}(c)-\bar{u}^{\text{mp}}(c)\}\qquad(13)$$

over $i=1,\ldots,N$, respectively. The result holds for arbitrary $c\in\mathbb{R}^J$.

Refer to (13) as the fully interacted aggregate specification for the missingness-pattern method. Theorem 2 ensures that $(\widehat{\tau}_{\text{L}}^{\text{mp}},\widehat{\text{se}}_{\text{L}}^{\text{mp}})$ are direct outputs of its OLS fit, with $u_i^{\text{mp}}(c)$ as the effective covariate vector analogous to the x_i^{ccov}, $x_i^{\text{imp}}(c)$, and $x_i^{\text{mim}}(c)$ in (4), (6), and (8), respectively. This ensures the equivalence of $\widehat{\tau}_{\text{L}}^{\text{mp}}$ and $\widehat{\tau}_{\text{L}}^{\text{mim}}$ when $J=1$ by Example 1 (continued) above, and allows us to quantify the unconditional asymptotic properties of $\widehat{\tau}_{\text{L}}^{\text{mp}}$ for general J.

Corollary 1. $u_i^{\text{mp}}(c)=x_i^{\text{mim}}(c)$ and $\widehat{\tau}_{\text{L}}^{\text{mp}}=\widehat{\tau}_{\text{L}}^{\text{mim}}$ for $J=1$.

Let v_{L}^{mp} and $S_{\tau\tau,\text{L}}^{\text{mp}}$ be the analogs of v_{L} and $S_{\tau,\text{L}}^2$ in Lemma 1 defined over $\{Y_i(0),Y_i(1),u_i^{\text{mp}}(0_J)\}_{i=1}^N$.

Proposition 5. Assume complete randomization and Condition S4 for $\{Y_i(0),Y_i(1),u_i^{\text{mp}}(0_J)\}_{i=1}^N$. Then $\sqrt{N}(\widehat{\tau}_{\text{L}}^{\text{mp}}-\tau)\rightsquigarrow\mathcal{N}(0,v_{\text{L}}^{\text{mp}})$, with $N(\widehat{\text{se}}_{\text{L}}^{\text{mp}})^2-v_{\text{L}}^{\text{mp}}=S_{\tau\tau,\text{L}}^{\text{mp}}+o_{\text{P}}(1)$, where $S_{\tau\tau,\text{L}}^{\text{mp}}\geq 0$.

Proposition 5 is a direct consequence of Lemma 1 and Theorem 2, and justifies the large-sample Wald-type inference based on $(\widehat{\tau}_{\text{L}}^{\text{mp}},\widehat{\text{se}}_{\text{L}}^{\text{mp}})$ irrespective of the missingness mechanism. The asymptotic efficiency of saturated model over its restricted variants further ensures the asymptotic efficiency of $\widehat{\tau}_{\text{L}}^{\text{mp}}$ over $\widehat{\tau}_{\text{F}}^{\text{mp}}$; see Theorem S2 in the supplementary materials.

Theorem 3 summarizes the relative efficiency between $\widehat{\tau}_{\text{L}}^{\dagger}$'s for the four consistent strategies, $\dagger\in\{\text{ccov},\text{imp},\text{mim},\text{mp}\}$. This, together with the efficiency of $\widehat{\tau}_{\text{L}}^{\dagger}$ over $\widehat{\tau}_{\text{F}}^{\dagger}$ for each individual strategy, ensures the asymptotic efficiency of $\widehat{\tau}_{\text{L}}^{\text{mp}}$ among all eight consistent estimators in Table 1.

Theorem 3. $v_{\text{L}}^{\text{mp}}\leq v_{\text{L}}^{\text{mim}}\leq v_{\text{L}}^{\text{imp}}(c_\infty)\leq v_{\text{L}}^{\text{ccov}}\leq v_{\text{N}}$.

Compare the definition of $u_i^{\text{mp}}(c)$ with $x_i^{\text{mim}}(c)$ to see that $u_i^{\text{mp}}(c)$ includes interaction terms like $x_i^{\text{imp}}(c)M_{ij}$, $x_i^{\text{imp}}(c)M_{ij}M_{ij'}$ that are not in $x_i^{\text{mim}}(c)$. This suggests the advantage of $\widehat{\tau}_{\text{L}}^{\text{mp}}$ over $\widehat{\tau}_{\text{L}}^{\text{mim}}$ when the covariates interact with the missingness pattern in affecting the treatment effect.

Despite the desired gain in large-sample efficiency, the missingness-pattern method can be demanding on the missingness-pattern-specific sample sizes in finite samples even with a moderate J. Denote by $J_{(m)}=\sum_{j=1}^J(1-m_j)$ the number of available covariates under missingness pattern m. The pattern-specific additive estimator $\widehat{\tau}_{\text{F},(m)}$ is well defined only if $N_{(m)}\geq J_{(m)}+2$; the pattern-specific fully interacted estimator $\widehat{\tau}_{\text{L},(m)}$ is well defined only if $\min\{N_{(m),0},N_{(m),1}\}\geq J_{(m)}+1$. When some $\widehat{\tau}_{*,(m)}$'s are not well defined due to these sample size constraints, we recommend going back to the missingness-indicator method to ensure finite sample feasibility.

5. Numerical Studies

5.1. Simulation

We use simulation to illustrate the finite sample properties of the proposed methods. The results are coherent with the asymptotic theory in Section 4, showing (i) the inconsistency of $\widehat{\tau}_*^{\text{cc}}$ $(*=\text{F},\text{L})$ when $\tau^{\text{cc}}\neq\tau$; (ii) the efficiency of $\widehat{\tau}_{\text{L}}^{\text{mim}}$ over $\{\widehat{\tau}_{\text{F}}^{\text{mim}},\widehat{\tau}_*^{\text{imp}}(c),\widehat{\tau}_*^{\text{ccov}}:*=\text{F},\text{L}\}$ when missingness is correlated with the potential outcomes; (iii) the efficiency of $\widehat{\tau}_{\text{L}}^{\text{mp}}$ over $\widehat{\tau}_{\text{L}}^{\text{mim}}$ in large samples when the potential outcomes depend on interactions between $\{M_{i1},\ldots,M_{iJ},x_i\}$, and (iv) the robustness of the proposed methods to model misspecification. We relegate the details to the supplementary materials due to space limitations.

5.2. Application

Angrist, Lang, and Oropoulos (2009) conducted a randomized field experiment in a Canadian university to evaluate the effect of academic services and incentives on academic performance. A random sample of 400 out of 1656 eligible first-year students was selected to receive the opportunity to win fellowships for meeting a target grade point average. One question of interest is how such fellowship opportunities affect the average grades in the fall semester.

Let Y_i and Z_i be the average fall grade and indicator of access to the fellowship opportunity for student i. A total of 1404 students have available fall grades, among which 338 were offered the fellowship opportunity and 1066 were not. We use

Table 2. Reanalysis of the data from Angrist, Lang, and Oropoulos (2009).

	$\widehat{\tau}_{\mathrm{F}}^{\dagger}$			$\widehat{\tau}_{\mathrm{L}}^{\dagger}$		
Strategy	Estimate	Robust s.e.	*p*-value	Estimate	Robust s.e.	*p*-value
cc	2.152	0.704	0.002	2.114	0.702	0.003
ccov	1.866	0.686	0.007	1.895	0.677	0.005
imp	1.906	0.686	0.006	1.876	0.680	0.006
mim	1.935	0.683	0.005	1.924	0.674	0.004
mp	1.998	0.681	0.003	1.926	0.675	0.004

NOTE: The outcome is the average fall grade, and the treatment is access to the fellowship opportunity. For comparison, $\widehat{\tau}_{\mathrm{N}} = 2.013$ and $\widehat{\mathrm{se}}_{\mathrm{N}} = 0.703$ with *p*-value = 0.004.

them as the study population, and follow Angrist, Lang, and Oropoulos (2009) to control for sex, mother tongue, high school grade quartile, number of credits enrolled, and responses to survey questions on procrastination and parents' education in the regression analysis. A total of 1278 students have all covariates observed, accounting for 91% of the study population. All incomplete cases are due to missing responses to all survey questions. This results in two missingness patterns, corresponding to the respondents and nonrespondents, respectively.

Table 2 summarizes the results from our reanalysis of the data. All five strategies yield coherent results about a significant effect of the fellowship opportunities on fall grades. The complete-case analysis yields the largest point estimates, 2.152 and 2.114, overall. The missingness-pattern-specific analysis of the incomplete cases, on the other hand, yields point estimates of 0.435 and 0.014 under the additive and fully interacted specifications, respectively. This suggests a high possibility of systematic differences between the respondents and nonrespondents, subjecting the complete-case analysis to possibly large biases.

Importantly, the reduction in the robust standard error of $\widehat{\tau}_{\mathrm{L}}^{\mathrm{mim}}$ in Table 2 can be an artifact of adding more covariates, and may not reflect the true variability in finite samples. To empirically investigate whether the differences in precision are coherent with the asymptotic theory, we simulate data based on Angrist, Lang, and Oropoulos (2009) to verify the finite sample efficiency of $\widehat{\tau}_{\mathrm{L}}^{\mathrm{mim}}$ in this example. We relegate the details to the supplementary materials.

6. Conclusion

We established the validity of the complete-covariate analysis, single imputation, the missingness-indicator method, and the missingness-pattern method for large-sample Wald-type inference regardless of (i) the relation between missingness, potential outcomes, and covariates, (ii) the correctness of the linear models, and (iii) the choice of the imputed values if any. Consistency is hence a rather basic criterion for evaluating regression-adjusted estimators, rendering the possible inconsistency of the complete-case analysis all the more undesirable.

Based on theory and simulation, we recommended using the missingness-indicator method along with Lin (2013)'s specification to adjust for missing covariates in randomized experiments. When the treatment does not affect the missingness indicators, the resulting estimator is consistent for the average treatment effect, and asymptotically more efficient than the unadjusted estimator and the estimators based on complete or imputed covariates alone. We also proposed the missingness-pattern method as a modification to reap additional asymptotic efficiency.

Due to space limitations, we relegate extensions to alternative regression specifications, cluster and stratified randomizations, the Fisher randomization test, rerandomization with missing covariates, and possible violations of Condition 1 to the supplementary materials.

Supplementary Materials

The supplementary materials contain additional results and technical details.

Acknowledgments

We thank Fan Li, Tianyu Guo, Georgia Papadogeorgous, and Zhi Geng for helpful discussions, and two Associate Editors and three referees for most constructive comments.

Funding

Zhao was funded by the Start-Up grant R-155-000-216-133 from the National University of Singapore. Ding was partially funded by the U.S. National Science Foundation # 1945136.

ORCID

Peng Ding http://orcid.org/0000-0002-2704-2353

References

Anderson, A. B., Basilevsky, A., and Hum, D. P. J. (1983), *Missing Data: A Review of the Literature*, volume 1 of Handbook of Survey Research, pp. 415–492, New York: Academic Press. [2]

Angrist, J., Lang, D., and Oropoulos, P. (2009), "Incentives and Services for College Achievement: Evidence from a Randomized Trial," *American Economic Journal: Applied Economics*, 1, 136–163. [9,10]

Berk, R., Pitkin, E., Brown, L., Buja, A., George, E., and Zhao, L. (2013), "Covariance Adjustments for the Analysis of Randomized Field Experiments," *Evaluation Review*, 37, 170–196. [3]

Carpenter, J. R., and Kenward, M. G. (2007), *Missing Data in Randomised Controlled Trials: A Practical Guide*. UK National Health Service, National Coordinating Centre for Research on Methodology. [2,4]

Chong, A., Cohen, I., Field, E., Nakasone, E., and Torero, M. (2016), "Iron Deficiency and Schooling Attainment in Peru," *American Economic Journal: Applied Economics*, 8, 222–255. [2]

Cohen, J., and Cohen, P. (1975), *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences*, New York: Lawrence Erlbaum Associates. [2,5]

D'Agostino, R. B., and Rubin, D. B. (2000), "Estimating and using Propensity Scores with Partially Missing Data," *Journal of the American Statistical Association*, 95, 749–759. [2]

Ding, P., and Geng, Z. (2014), "Identifiability of subgroup causal effects in randomized experiments with nonignorable missing covariates," *Statistics in Medicine*, 33, 1121–1133. [2]

Donders, A. R. T., van der Heijden, G. J., Stijnen, T., and Moons, K. G. M. (2006), "Review: A Gentle Introduction to Imputation of Missing Values," *Journal of Clinical Epidemiology*, 59, 1087–1091. [2]

Fisher, R. A. (1935), *The Design of Experiments* (1st ed.), Edinburgh, London: Oliver and Boyd. [1,3]

Fogarty, C. B., Mikkelsen, M. E., Gaieski, D. F., and Small, D. S. (2016), "Discrete Optimization for Interpretable Study Populations and Randomization Inference in an Observational Study of Severe Sepsis Mortality," *Journal of the American Statistical Association*, 111, 447–458. [2]

Freedman, D. A. (2008), "On Regression Adjustments to Experimental Data," *Advances in Applied Mathematics*, 40, 180–193. [2,3]

Gerber, A. S., and Green, D. P. (2012), *Field Experiments: Design, Analysis, and Interpretation*, New York: W. W. Norton and Company. [2]

Glasser, M. (1964), "Linear Regression Analysis with Missing Observations among the Independent Variables," *Journal of the American Statistical Association*, 59, 834–844. [2]

Greenland, S., and Finkle, W. D. (1995), "A Critical Look at Methods for Handling Missing Covariates in Epidemiologic Regression Analyses," *American Journal of Epidemiology*, 142, 1255–1264. [2]

Groenwold, R. H., White, I. R., Donders, A. R., Carpenter, J. R., Altman, D. G., and Moons, K. G. (2012), "Missing Covariate Data in Clinical Research: When and when not to Use the Missing-Indicator Method for Analysis," *Canadian Medical Association Journal*, 184, 1265–1269. [2]

Haitovsky, Y. (1968), "Missing Data in Regression Analysis," *Journal of the Royal Statistical Society*, Series B, 30, 67–82. [2]

Ibrahim, J. G., Chen, M.-H., Lipsitz, S. R., and Herring, A. H. (2005), "Missing-Data Methods for Generalized Linear Models: A Comparative Review," *Journal of the American Statistical Association*, 100, 332–346. [1,2]

Imbens, G. W., and Rubin, D. B. (2015), *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction*, Cambridge: Cambridge University Press. [2,3]

Jones, M. P. (1996), "Indicator and Stratification Methods for Missing Explanatory Variables in Multiple Linear Regression," *Journal of the American Statistical Association*, 91, 222–230. [2,8]

Kamat, G., and Reiter, J. P. (2021), "Leveraging Random Assignment to Impute Missing Covariates in Causal Studies," *Journal of Statistical Computation and Simulation*, 91, 1275–1305. [5]

Kayembe, M. T., Jolani, S., Tan, F. E. S., and van Breukelen, G. J. P. (2020), "Imputation of Missing Covariate in Randomized Controlled Trials with a Continuous Outcome: Scoping Review and New Results," *Pharmaceutical Statistics*, 19, 840–860. [5]

Li, X., and Ding, P. (2020), "Rerandomization and Regression Adjustment," *Journal of the Royal Statistical Society*, Series B, 82, 241–268. [3]

Lin, W. (2013), "Agnostic Notes on Regression Adjustments to Experimental Data: Reexamining Freedman's Critique," *Annals of Applied Statistics*, 7, 295–318. [1,2,3,10]

Little, R. J. A. (1992), "Regression with Missing x's: A Review," *Journal of the American Statistical Association*, 87, 1227–1237. [1,2,5]

——— (1993), "Pattern-Mixture Models for Multivariate Incomplete Data," *Journal of the American Statistical Association*, 88, 125–134. [6]

Little, R. J. A., and Rubin, D. B. (2019), *Statistical Analysis with Missing Data* (3rd ed.), New York: Wiley. [1,2]

Mattei, A. (2009), "Estimating and using Propensity Score in Presence of Missing Background Data: An Application to Assess the Impact of Childbearing on Wellbeing," *Statistical Methods and Applications*, 18, 257–273. [2]

Matthai, A. (1951), "Estimation of Parameters from Incomplete Data with Application to Design of Sample Surveys," *Sankhya*, 11, 145–152. [2,6]

Miettinen, O. S. (1985), *Theoretical Epidemiology: Principles of Occurrence Research in Medicine*, New York: Wiley. [2]

Miratrix, L. W., Sekhon, J. S., and Yu, B. (2013), "Adjusting Treatment Effect Estimates by Post-Stratification in Randomized Experiments," *Journal of the Royal Statistical Society*, Series B, 75, 369–396. [2,6,8]

Negi, A., and Wooldridge, J. M. (2021), "Revisiting Regression Adjustment in Experiments with Heterogeneous Treatment Effects," *Econometric Reviews*, 40, 504–534. [3]

Neyman, J. (1923), "On the Application of Probability Theory to Agricultural Experiments," (with Discussion), *Statistical Science*, 5, 465–472. [1,2,3]

Robins, J. M., Rotnitzky, A., and Zhao, L. P. (1994), "Estimation of Regression Coefficients when some Regressors are not Always Observed," *Journal of the American Statistical Association*, 89, 846–866. [2]

Rosenbaum, P. R. (2010), *Design of Observational Studies*, New York: Springer. [2]

Rosenbaum, P. R., and Rubin, D. B. (1984), "Reducing Bias in Observational Studies using Subclassification on the Propensity Score," *Journal of the American Statistical Association*, 79, 516–524. [2,6]

Rubin, D. B. (1976), "Inference and Missing Data," *Biometrika*, 63, 581–592. [2,7]

——— (1987), *Multiple Imputation for Nonresponse in Surveys*, New York: Wiley. [1,2,5]

Schemper, M., and Smith, T. L. (1990), "Efficient Evaluation of Treatment Effects in the Presence of Missing Covariate Values," *Statistics in Medicine*, 9, 777–784. [5,7]

Sullivan, T. R., White, I. R., Salter, A. B., Ryan, P., and Lee, K. J. (2018), "Should Multiple Imputation be the Method of Choice for Handling Missing Data in Randomized Trials?" *Statistical Methods in Medical Research*, 27, 2610–2626. [4,5]

White, I. R., and Carlin, J. B. (2010), "Bias and Efficiency of Multiple Imputation Compared with Complete-Case Analysis for Missing Covariate Values," *Statistics in Medicine*, 29, 2920–2931. [1]

White, I. R., and Thompson, S. G. (2005), "Adjusting for Partially Missing Baseline Measurements in Randomized Trials," *Statistics in Medicine*, 24, 993–1007. [2,4]

Wilks, S. S. (1932), "Moments and Distributions of Estimates of Population Parameters from Fragmentary Samples," *Annals of Mathematical Statistics*, 3, 163–195. [2,6]

Yang, S., Wang, L., and Ding, P. (2019), "Causal Inference with Confounders Missing not at Random," *Biometrika*, 106, 875–888. [2]

Ye, T., Shao, J., Yi, Y. Y., and Zhao, Q. Y. (2022), "Toward Better Practice of Covariate Adjustment in Analyzing Randomized Clinical Trials," *Journal of the American Statistical Association*, DOI:10.1080/01621459.2022.2049278. [3]

Zhao, A., and Ding, P. (2021), "Covariate-Adjusted Fisher Randomization Tests for the Average Treatment Effect," *Journal of Econometrics*, 225, 278–294. [3]