Original Research Article

SMMR
STATISTICAL METHODS IN MEDICAL RESEARCH

Statistical Methods in Medical Research
1–14

DOI: 10.1177/09622802221135251
journals.sagepub.com/home/smm
SAGE

Sensitivity analyses in longitudinal clinical trials via distributional imputation

Siyi Liu[1], Shu Yang[1], Yilong Zhang[2], and Guanghan (Frank) Liu[2]

Abstract

Missing data is inevitable in longitudinal clinical trials. Conventionally, the missing at random assumption is assumed to handle missingness, which however is unverifiable empirically. Thus, sensitivity analyses are critically important to assess the robustness of the study conclusions against untestable assumptions. Toward this end, regulatory agencies and the pharmaceutical industry use sensitivity models such as return-to-baseline, control-based, and washout imputation, following the ICH E9(R1) guidance. Multiple imputation is popular in sensitivity analyses; however, it may be inefficient and result in an unsatisfying interval estimation by Rubin's combining rule. We propose distributional imputation in sensitivity analysis, which imputes each missing value by samples from its target imputation model given the observed data. Drawn on the idea of Monte Carlo integration, the distributional imputation estimator solves the mean estimating equations of the imputed dataset. It is fully efficient with theoretical guarantees. Moreover, we propose weighted bootstrap to obtain a consistent variance estimator, taking into account the variabilities due to model parameter estimation and target parameter estimation. The superiority of the distributional imputation framework is validated in the simulation study and an antidepressant longitudinal clinical trial.

Keywords

Longitudinal clinical trial, missing data, distributional imputation, multiple imputation, sensitivity analysis

1 Introduction

In longitudinal clinical trials, participants are likely to deviate from the protocol that causes the missing data. The deviations from the protocol may include poor compliance with the treatment or loss of follow-ups. In the presence of missing data, the importance of defining an appropriate estimand has been put forward by the ICH E9(R1) working group.[1] Following the instructions, the estimand should give a precise description of the treatment effect of interest from a population perspective, and account for the intercurrent events such as the discontinuation of treatment.

To take into account missingness for a proper estimator of an estimand, Rubin[2] proposes three missing mechanisms as missing completely at random (MCAR), missing at random (MAR), and missing not at random (MNAR). Missingness that is not related to any components of the data, for example, participants dropping out of the trial due to work or family considerations, is categorized as MCAR. While in most clinical studies involving patients with missing outcomes, it is likely that the missingness depends on the health status of patients. For example, individuals with severe outcomes are more likely to drop out from the study or switch to certain rescue therapies. MAR is typically used in longitudinal clinical trials targeting the primary analysis, which assumes the same conditional outcome distribution between the participants who remain in the study and the ones who drop out, given the baseline covariates and the historical outcomes,[3] that is, the participants are expected to take the assigned treatment even after the occurrence of missingness. However, the

[1]Department of Statistics, North Carolina State University, Raleigh, NC, USA
[2]Merck & Co., Inc., Kenilworth, NJ, USA

Corresponding author:
Shu Yang, Department of Statistics, North Carolina State University, Raleigh, NC, USA.
Email: syang24@ncsu.edu

MAR assumption is not verifiable and may be violated for some drugs with a short half-life, where the treatment effect quickly fades away once the individuals discontinue from the active treatment, leading to a MNAR assumption. Therefore, it is vital to conduct sensitivity or supplemental analyses to explore the robustness of results to alternative MNAR-related assumptions as recommended by the US Food and Drug Administration (FDA) and National Research Council.[4] Throughout the paper, we evaluate the treatment effect under scenarios that deviate from MAR and call these settings sensitivity analyses for simplicity.

Literature has covered a wide range of methods to handle missingness in sensitivity analyses. Based on the selection model framework, one can introduce the external sensitivity parameter to model the missingness indicator.[5–7]Alternatively, Steenland and Greenland[8] posit a prior distribution of the sensitivity parameter with the incorporation of Monte Carlo (MC) sampling from the Bayesian perspective. Under the pattern-mixture model (PMM) framework,[9] sensitivity analyses can be formulated by specifying the distributions of missing outcomes in each missingness pattern.[10,11] In the paper, we focus on the PMM framework and consider several plausible missingness scenarios under MNAR, which we call the "sensitivity models." Our main focus in this paper is on the jump-to-reference (J2R) scenario proposed by Carpenter et al.,[3] which assumes that the missing outcomes in both treatments will have the same distributional profile as those in the control group with the same covariates. Its plausibility reveals under the sensitivity setting where the effect of the test drug is limited to a short period of time after the individual drops out. We also briefly introduce other sensitivity models such as return-to-baseline (RTB) and washout imputation, which have been used in the FDA reports for certain treatments.[12] While the RTB model expects a return to the baseline status for the dropouts yet may not show conservativeness if the missing data is imbalanced between the treatment groups, the washout imputation model combines the idea of J2R and RTB to overcome this imbalance issue. Although we focus on specific sensitivity models, our framework is extendable to other imputation mechanisms and the mixture of imputation strategies in sensitivity analyses.

Among the extensive methodologies in sensitivity analyses, the likelihood-based method and multiple imputation (MI) are the two most common approaches under the PMM framework. The likelihood-based method typically utilizes the ignorability of the missing mechanism under MAR to draw valid maximum-likelihood inferences given variation independence, that is, the parameters that control the missing mechanism and the model parameters are separable. For longitudinal clinical trials with continuous responses, one can fit a mixed model with repeated measures (MMRM) and incorporate the missing information to obtain inferences.[13,14] While it is efficient, the analytical form for the treatment effect estimator is only feasible to derive under restrictive scenarios such as normality or when dealing with mean types of estimands; and it requires repeats of tedious calculations if the missingness pattern changes. MI developed by Rubin[15] resorts to using computational techniques to ease the analytical requirements from the likelihood-based method. The FDA and National Research Council[4] highly recommend the use of MI and Rubin's MI combining rules to get inferences due to their flexibility and simplicity. Cro et al.[11] provide practical guidance on implementing MI in sensitivity analyses in longitudinal clinical studies. However, the severe inefficiency of the MI estimator in terms of interval estimation has been detected in the literature. Meng[16] discovers the inconsistency of the MI variance estimator when congeniality is not satisfied, that is, the imputation and analysis models contain different sources of information. Even when the imputation and analysis models are the same and correctly specified, Robins and Wang[17] find that the variance estimation using Rubin's rule may not be consistent. Although the consistency of Rubin's variance estimator generally holds under MAR in the primary analysis, the overestimation of the variance using Rubin's rule is frequently detected in sensitivity analyses.[18,19] The motivating example in Section 2 further shows an alteration of the study conclusion due to the conservative variance estimate, where the same statistically significant treatment effect fails to be detected in the sensitivity analysis, rising a dilemma for the investigators in the process of decision-making.

In this paper, we propose distributional imputation (DI) based on the idea of MC integration[20] and develop a unified framework to conduct sensitivity analyses using DI in longitudinal clinical trials. The motivation of DI is to impute the missing components from the target imputation model given the observed data and use the mean estimating equations approximated by MC integration to draw efficient inferences. Moreover, we put forward a weighted bootstrap procedure for variance estimation, which incorporates the uncertainty from the model parameter estimation and the target parameter estimation. The DI estimator drawn from our framework is fully efficient with the firm theoretical ground, and the weighted-bootstrap variance estimator is consistent with straightforward realization and the avoidance of re-imputing the missing components compared to the conventional bootstrap methods.

The rest of the paper proceeds as follows. Section 2 uses antidepressant clinical trial data to motivate the development of an efficient imputation method. Section 3 introduces the basic setup, provides notations, estimands, imputation mechanisms in sensitivity analyses, and comments on existing methods to handle missingness. Section 4 presents DI and its main steps. Section 5 gives the asymptotic theories for the DI estimator and proposes weighted bootstrap on variance estimation. Section 6 explores the finite-sample performance of the DI estimator via simulation. Section 7 applies

the proposed framework to the motivating example. Section 8 draws the conclusion. Supplemental Material contains the technical setup, proof of the theorems, additional simulation and real-data application results, and codes for the DI implementation.

2 Motivating example

An antidepressant clinical trial from the Auspices of the Drug Information Association is conducted to evaluate the effect of an experimental medication.[21] The study measures the longitudinal outcomes of the Hamilton Depression Rating Scale for 17 items (HAMD-17) at baseline and weeks 1, 2, 4, 6, and 8 for 200 patients who were randomly assigned to the control and treatment groups at a 1:1 ratio. We use the fully observed baseline responses as the baseline covariates. The post-baseline measurements in the data reflect a typical missing data issue in longitudinal clinical trials, where 39 patients in the control group and 30 patients in the treatment group dropped out during the study period, and one patient in the treatment group contains intermittent missingness. For illustration purposes, we assume MAR for this intermittent missingness and focus on the data with monotone missingness. The relatively high dropout rate prompts the need for imputation to utilize the information related to the missingness.

To investigate the treatment effect in different aspects, we explore two population-level summaries by constructing different estimands for treatment evaluation. The first treatment effect estimand is the average treatment effect (ATE) defined by the mean difference between the relative change of the HAMD-17 score from the baseline value at the last visit. The second estimand is the risk difference defined by the percentage difference of patients with 50% or more improvement from the baseline HAMD-17 score at the last visit.

A scrutiny of the data in Section S3.3 in the Supplemental Material indicates an underlying normal outcome distribution. Under the normal assumption, we conduct MI to the missing components with the imputation size as 100 under MAR to perform the primary analysis and under J2R for the sensitivity analysis, and we obtain the inferences using Rubin's rule. The point estimates of the ATE and the risk difference accompanied with the 95% CIs are presented in Figure 1. In the primary analysis that assumes MAR, both the ATE and the risk difference reveal a significant treatment effect. However, the sensitivity analysis under J2R fails to capture the same significance under MI, leading to a loss of the credibility of the experimental drug and a potential influence on the decision made by the investigators. The alteration of the study conclusion may result from the overestimation issue of Rubin's MI variance estimator as detected in the literature involving sensitivity analyses,[19] rather than the loss of effectiveness in the test drug. To further explore the cause of the altered study result, it is vital to overcome the overestimation issue brought up from MI and develop an efficient imputation approach to obtain a consistent variance estimator without an expensive computational cost.

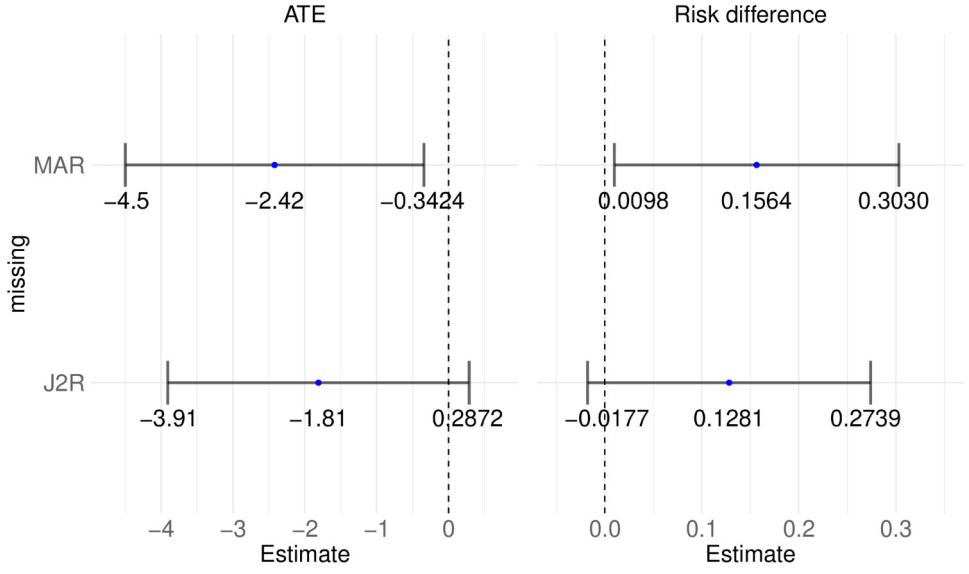

Figure 1. Estimation results of the ATE and the risk difference under MAR and J2R, accompanied with the 95% CIs. ATE: average treatment effect; MAR: missing at random; J2R: jump-to-reference; 95% CIs: 95% confidence intervals.

3 Basic setup

3.1 Notations and estimands

Let Y_{ik} be the continuous response for patient i at time t_k, where $i = 1, \ldots, n$, and $k = 1, \ldots, T$. Denote the baseline p-dimensional fully observed covariate vector as X_i, the group indicator as G_i ranging from 1 to J to represent J distinct treatment groups, and the observed indicator as $R_i = (R_{i1}, \ldots, R_{iT})^\intercal$ for patient i, where $R_{ik} = 1$ if Y_{ik} is observed, $R_{ik} = 0$ otherwise. Without loss of generality, we consider $J = 2$, where $G_i = 1, 2$ represents the ith patient is in the control or active treatment group, respectively. Let $Y_i = (Y_{i1}, \ldots, Y_{iT})^\intercal$ be a T-dimensional longitudinal vector containing history and current information. A monotone missingness pattern is assumed, that is, if missingness begins at time t^*, we have $R_{ik} = 1$ for $t_k < t^*$ and $R_{ik} = 0$ for $t_k \geq t^*$. We can partition each response as $Y_i = (Y_{\text{obs},i}, Y_{\text{mis},i})^\intercal$, where $Y_{\text{obs},i}$ and $Y_{\text{mis},i}$ are the observed and missing components. Denote $Z_i = (Y_i, X_i, G_i, R_i)$ as the full data for patient i. In the presence of missing data, denote $Z_{\text{obs},i} = (Y_{\text{obs},i}, X_i, G_i, R_i)$ as the observed data. Then, $Z_i = (Z_{\text{obs},i}, Z_{\text{mis},i})$ corresponds to the combination of the observed and missing parts.

For the treatment comparison, we consider different treatment effect estimands based on different population-level summaries defined through the estimating equations as $\tau = \tau_2 - \tau_1$, where τ_j is the value such that $\mathbb{E}\{\psi_j(Z_i, \tau_j)\} = 0$ and $\psi_j(Z, \tau_j)$ is a function in the space $\mathbb{R}^d \times \Omega$, with Ω as a compact subset of the Euclidean space and $\psi_j(Z, \tau_j)$ as a continuous function of τ_j for each vector Z and measurable of Z for each τ_j.

The expectation $\mathbb{E}\{\psi_j(Z_i, \tau_j)\}$ prompts the determination of the full-data distribution. Under MNAR, we use the PMM framework to describe the data distribution as $f(Z) = \int f(Z \mid R) f(R) dR$, where $Z_1, \ldots, Z_n$ are independently sampled from a parametric model $f(Z, \theta_0)$ with the support free of θ_0. Here, $Z \in \mathbb{R}^d$ is a d-dimensional vector, Θ is a q-dimensional Euclidean space, and θ_0 is the true model parameter lying in the interior of Θ. Moreover, let $s(Z, \theta)$ be the score function of $f(Z, \theta)$, and assume $s(Z, \theta)$ to be a continuous function of θ for each Z and a measurable function of Z for each θ. The identification of the treatment effect relies on the assumption of the pattern-specific data distribution $f(Z \mid R)$ under each missingness pattern, which is prespecified in a statistical analysis plan for clinical trials as the sensitivity models under hypothetical scenarios to address potential intercurrent events that may impact the estimation of the treatment effect.[1]

The most popular full-data model in longitudinal clinical trials is the MMRM as recommended by the FDA and National Research Council,[4] which assumes an underlying multivariate normal distribution for the longitudinal outcomes. The motivating example in Section 2 validates its application in practice. Therefore, throughout the paper, we assume the continuous longitudinal outcomes Y_i given the covariates X_i and the group indicator $G_i = j$ independently follow a multivariate normal distribution as

$$Y_i \mid (X_i, G_i = j) \sim \mathcal{N}_T(\mu_{ij}, \Sigma^{(j)}) \tag{1}$$

where $\mu_{ij} = (\mu_{ij1}, \ldots, \mu_{ijT})^\intercal = (\tilde{X}_i^\intercal \beta_{j1}, \ldots, \tilde{X}_i^\intercal \beta_{jT})^\intercal$, $\beta_{j1}, \ldots, \beta_{jT}$ are $(p+1)$-dimensional group-specific vectors, $\tilde{X}_i = (1, X_i^\intercal)^\intercal$, and $\Sigma^{(j)}$ is a group-specific covariance matrix.

When no missingness is involved in the data, the only pattern corresponds to $R = \mathbf{1}_T$, where $\mathbf{1}_T$ is a T-dimensional all-ones vector representing the outcomes are fully observed. The treatment effect identification boils down to the specification of the conditional distribution $Y_i \mid (X_i, G_i = j)$, which has been specified in the formula (1). Under this circumstance, we present three typical population-level summaries to define the treatment effect estimands in the following example.

Example 1. (Treatment effect estimands)The parameter $\tau = \tau_2 - \tau_1$ can represent different types of the treatment effect given the following choices of $\psi_j(\cdot)$ for $j = 1, 2$:

1. The ATE when $\tau_j = \mathbb{E}(Y_{iT} \mid G_i = j)$: $\psi_j(Z_i, \tau_j) = I(G_i = j)(Y_{iT} - \tau_j)$, where $I(\cdot)$ is the indicator function.
2. The risk difference when $\tau_j = P(Y_{iT} \geq c \mid G_i = j)$: $\psi_j(Z_i, \tau_j) = I(G_i = j)\{I(Y_{iT} \geq c) - \tau_j\}$, where c is a prespecified threshold.
3. Distributional information of the treatment, for example, the quantile treatment effect (QTE) for the qth quantile of responses when $\tau_{j,q}$ is the qth quantile of the outcome distribution at the last time point: $\psi_j(Z_i, \tau_{j,q}) = I(G_i = j)\{I(Y_{iT} \leq \tau_{j,q}) - q\}$.

Among the above estimands, the ATE is most widely used to describe the treatment effect in clinical trials.[3,14,19] Some clinical studies also care about the risk difference regarding the percentage of patients with the endpoint continuous outcomes dichotomized by a certain threshold for each group. For example, Roussel et al.[22] take the difference of the percentage of participants with HbA1c <7% in each group as a secondary endpoint. However, the ATE is possibly insufficient to

capture the treatment effect under a skewed outcome distribution, where the test drug may not influence the average outcome, but the tail of the outcome distribution. In these cases, the QTE is preferred.[23]

3.2 Sensitivity models in sensitivity analyses

When there is more than one missingness pattern, the identification of the treatment effect is accomplished by specifying the data distribution $f(Z \mid R)$ under each pattern. Since the distribution $f(Z \mid R)$ is unobserved if $R \neq \mathbf{1}_T$, several plausible sensitivity models are proposed to model the missingness. The ICH E9(R1) addendum[1] addresses the importance of specifying explicit MNAR assumptions that underlie the sensitivity analysis in advance of the clinical trial based on different characteristics of drugs. Here, we concentrate on the J2R assumption to assess the robustness of the study conclusions.

The J2R sensitivity model is one specific control-based imputation model, which envisions the missing responses in the treatment group will have the same outcome profile as those in the control group with the same baseline covariates after dropout.[3] It can be a conservative missingness assumption that assumes the treatment effect disappears immediately after patients discontinue the active treatment, and it is a commonly used sensitivity analysis in practice for drugs with a short half-life.[21,24] If we are able to acquire additional knowledge regarding the missing components, the imputation scheme can be readily formulated based on those post-deviation data. While such extra information is often unavailable, a normal distribution conditional on the observed components is typically assumed to construct the imputation model in sensitivity analyses.[11] Equipped with the normality assumption, the group-specific model for the missing outcomes is the conditional distribution given the observed data under each missingness pattern as

$$\begin{aligned} &Y_{\text{mis},i} \mid (Y_{\text{obs},i}, X_i, G_i = j, R_{ik-1} = 1, R_{ik} = 0) \sim \\ &\quad \mathcal{N}_{T-k}\{\mu^{(k)}_{\text{mis},ij} + \Sigma^{(1)}_{21}\Sigma^{(1)-1}_{11}(Y_{\text{obs},i} - \mu^{(k)}_{\text{obs},ij}), \Sigma^{(1)}_{22} - \Sigma^{(1)}_{21}\Sigma^{(1)-1}_{11}\Sigma^{(1)}_{12}\} \end{aligned} \tag{2}$$

where $\mu^{(k)}_{\text{mis},ij}, \mu^{(k)}_{\text{obs},ij}$ are the individual-specific mean vectors and $\Sigma^{(1)} = \begin{pmatrix} \Sigma^{(1)}_{11} & \Sigma^{(1)}_{12} \\ \Sigma^{(1)}_{21} & \Sigma^{(1)}_{22} \end{pmatrix}$ is the covariance matrix for the control group partitioned corresponding to $Y_{\text{obs},i}$ and $Y_{\text{mis},i}$. Therefore, modeling the J2R sensitivity model corresponds to specifying the group-specific mean vector and covariance.

Assumption 1. (Jump-to-reference model) For the control group, the missing components are MAR. The imputation model is of the form (2), with $\mu^{(k)}_{\text{mis},i1} = (\tilde{X}_i^{\mathsf{T}}\beta_{1(k+1)}, \ldots, \tilde{X}_i^{\mathsf{T}}\beta_{1T})^{\mathsf{T}}$ and $\mu^{(k)}_{\text{obs},i1} = (\tilde{X}_i^{\mathsf{T}}\beta_{11}, \ldots, \tilde{X}_i^{\mathsf{T}}\beta_{1k})^{\mathsf{T}}$.

For the treatment group, the imputation model is of the form (2), with $\mu^{(k)}_{\text{mis},i2} = (\tilde{X}_i^{\mathsf{T}}\beta_{1(k+1)}, \ldots, \tilde{X}_i^{\mathsf{T}}\beta_{1T})^{\mathsf{T}}$ and $\mu^{(k)}_{\text{obs},i2} = (\tilde{X}_i^{\mathsf{T}}\beta_{21}, \ldots, \tilde{X}_i^{\mathsf{T}}\beta_{2k})^{\mathsf{T}}$ representing the regression coefficients for the participants after deviation will "jump to" the ones in the control group with the same baseline covariates.

Apart from J2R as a way to quantify the deviation from MAR, we also summarize two more MNAR assumptions as the RTB and washout sensitivity models used in the FDA reports[12] to represent different ways to model the missing outcomes regardless of compliance after discontinuation in sensitivity analyses. The RTB model assumes a washout effect for the missing responses at the last time point in both groups, indicating that the outcome of interest will return to the baseline performance regardless of the prior treatment after dropping out. However, the biological plausibility of the washout assumption needs to be carefully evaluated, and RTB is not necessarily conservative when missing data is imbalanced between the control and treatment groups.[25]

Assumption 2. (Return-to-baseline model) The imputation model of the outcomes at the last time point follows the marginal baseline model $Y_{iT} \mid (X_i, R_{iT} = 0, G_i = j) \sim \mathcal{N}(\mu_{ij1}, \Sigma^{(j)}_{(1,1)})$, where $\Sigma^{(j)}_{(1,1)}$ represents the (1, 1) element of $\Sigma^{(j)}$.

The washout model also acts as a possible sensitivity model and appears in several statistical reports.[26] It combines the idea of the RTB and J2R assumptions by assuming a MAR pattern for the control group and an RTB pattern for the missing outcomes in the treatment group. One of the reasons to consider washout imputation is to address the potential issue of imbalanced missing data in RTB.

Assumption 3. (Washout model) For the control group, the assumption for the missing responses is MAR. The model is the same as the one for the control group in Assumption 1. For the treatment group, the assumption for the missing responses is the same as Assumption 2.

Given a prespecified sensitivity model that characterizes the MNAR assumption, one can therefore determine the pattern-specific data distribution $f(Z \mid R)$ and identify the treatment effect under the PMM framework. To obtain valid treatment effect inferences, one can implement the conventional likelihood-based or imputation approach to deal with missingness in sensitivity analyses. Both methods are elaborated in the next section.

3.3 Existing methods to handle missingness in sensitivity analyses

The likelihood-based method and MI are two traditional approaches to handle missingness in sensitivity analyses in longitudinal clinical trials. The likelihood-based method utilizes the MMRM model and the ignorability of the missing components under MAR to draw valid inferences. In terms of the MNAR-related sensitivity models, the analytical form of the inference is obtained via averaging over the dropout patterns based on the PMM framework.[19,13] However, the treatment effect estimator can be infeasible to derive under cases where the normality assumption is violated or the estimand of interest is not of the mean type. For example, when we focus on the risk difference of the test drug in the antidepressant trial in Section 2, complexity arises when incorporating the dropout patterns. Moreover, the likelihood-based estimator needs to be re-derived for different imputation mechanisms and can be complicated if there are multiple missingness patterns.

MI provides a simple way to handle diverse types of estimands. It creates multiple complete datasets by conducting imputations based on the prespecified imputation model and has Rubin's combining rule to obtain inferences. We illustrate one typical strategy that has often appeared in the literature[3,27] for conducting MI with Rubin's rule, in the sensitivity analysis under J2R in longitudinal clinical trials using the estimands in Example 1 as follows.

Step 1. For the observed data in the control group, fit the MMRM and obtain the estimated sensitivity model.
Step 2. Impute the missing values in both groups from the sensitivity model specified in Assumption 1. Repeat the imputation M times to create M imputed datasets.
Step 3. For each imputed dataset, conduct a complete data analysis by solving the estimating equations that correspond to the estimands in Example 1 and obtain $\hat{\tau}_{\mathrm{MI}}^{(m)}$ as the estimator of the mth imputed dataset, where $m = 1, \ldots, M$.
Step 4. Combine the estimations from the M imputed datasets by Rubin's combining rule and obtain the MI estimator as $\hat{\tau}_{\mathrm{MI}} = M^{-1} \sum_{m=1}^{M} \hat{\tau}_{\mathrm{MI}}^{(m)}$, with the variance estimator

$$\hat{\mathbb{V}}(\hat{\tau}_{\mathrm{MI}}) = \frac{1}{M} \sum_{m=1}^{M} \hat{\mathbb{V}}(\hat{\tau}_{\mathrm{MI}}^{(m)}) + \left(1 + \frac{1}{M}\right) B_M$$

where $B_M = (M-1)^{-1} \sum_{m=1}^{M} (\hat{\tau}_{\mathrm{MI}}^{(m)} - \hat{\tau}_{\mathrm{MI}})^2$ represents the between-imputation variance.

Although MI in conjunction with Rubin's rule acts as a standard approach in the analysis of clinical trials[4] and shows satisfying performance under MAR in the primary analysis, the overestimation issue of Rubin's variance estimator under the MNAR-related sensitivity models has frequently been addressed in the literature.[18,19,28] For example, Liu and Pang[19] find that the variance estimator using Rubin's rule tends to overestimate the true variance in simulation studies under J2R. The motivating example in Section 2 further captures a change in the study result using MI with Rubin's rule under J2R in the sensitivity analysis, which may result from the overestimation issue.

Therefore, a more efficient method to get valid estimators for diverse types of the treatment effect estimands and the corresponding appropriate variance estimators with a simple implementation is needed. We propose DI based on the idea of MC integration to get the inference and the weighted bootstrap procedure to obtain a consistent variance estimator.

4 Distributional imputation

We establish the DI framework to evaluate the treatment effect in sensitivity analyses. Given the parametric distributions of the missing components based on certain sensitivity models, the key insight is to impute each missing value by samples from its conditional distribution given the observed data. Drawn on the idea of MC integration, any estimating equations applied to the imputed dataset approximate the mean estimating equations given the observed data and thus allow an efficient estimation of the target estimand.

The use of the mean estimating equations conditional on the observed data to assess the treatment effect is prevalent in the missing data literature. Louis[29] takes advantage of the conditional mean estimating equations with the expectation-maximization algorithm to obtain valid inferences for the incomplete data. Robins and Wang[17] also apply the idea of the mean estimating equations to allow for the incompatibility between the imputation and analysis model. In the presence

of missing data, one can estimate the function $\psi_j(Z_i, \tau_j)$ that characterizes the treatment effect by the conditional expectation given the observed values under certain sensitivity models that have been prespecified in the trial protocol. Therefore, a consistent estimator of τ_j for $j = 1, 2$ is the solution to

$$\sum_{i=1}^{n} \mathbb{E}\{\psi_j(Z_i, \tau_j) \mid Z_{\text{obs},i}, \hat{\theta}\} = 0 \tag{3}$$

where $\hat{\theta}$ is a consistent estimator of an unknown modeling parameter $\theta \in \Theta$. A common choice of $\hat{\theta}$ is the pseudo maximum likelihood estimator given the observed data, that is, it solves the mean score equations

$$\frac{1}{n}\sum_{i=1}^{n} \mathbb{E}\{s(Z_i, \theta) \mid Z_{\text{obs},i}\} = 0 \tag{4}$$

Note that the mean estimating equations in (3) and (4) have general forms which can accommodate different sensitivity models. In longitudinal clinical trials, the typical mean estimating equations correspond to the score function of the MMRM for the observed data.

Even under the multivariate normal assumption, the explicit form of the target estimator $\hat{\tau}_j$ is only feasible to obtain when the function $\psi_j(Z_i, \tau_j)$ has a linear form such as the one in Example 1(a). However, we can estimate the conditional expectation $\mathbb{E}\{\psi_j(Z_i, \tau_j) \mid Z_{\text{obs},i}, \hat{\theta}\}$ using the complete data after imputation. For the missing component of ith continuous response Y_i, we independently draw $Y_{\text{mis},i}^{*(1)}, \ldots, Y_{\text{mis},i}^{*(M)}$ from a prespecified sensitivity model with the estimated conditional distribution $f(Z_{\text{mis},i}^{*(m)} \mid Z_{\text{obs},i}, \hat{\theta})$ such as Assumptions 1 to 3 used in sensitivity analyses.

With the imputed data, denote $Y_i^{*(m)} = (Y_{\text{obs},i}, Y_{\text{mis},i}^{*(m)})$ as the imputed longitudinal responses and $Z_i^{*(m)} = (Z_{\text{obs},i}, Z_{\text{mis},i}^{*(m)})$ as the full imputed data. DI incorporates the idea of MC integration. When the imputation size M is large, one can estimate the conditional expectation in (3) as

$$\mathbb{E}\{\psi_j(Z_i, \tau_j) \mid Z_{\text{obs},i}, \hat{\theta}\} \approx \frac{1}{M}\sum_{m=1}^{M} \psi_j(Z_i^{*(m)}, \tau_j) \tag{5}$$

Based on the MC approximation, we can therefore derive the DI estimator $\hat{\tau}_{\text{DI},j}$ for jth group by solving the estimating equations

$$\frac{1}{M}\sum_{i=1}^{n}\sum_{m=1}^{M} \psi_j(Z_i^{*(m)}, \tau_j) = 0 \tag{6}$$

Example 2. (DI estimator for the treatment effect) For all estimands in Example 1, the DI estimator of the treatment effect is $\hat{\tau}_{\text{DI}} = \hat{\tau}_{\text{DI},2} - \hat{\tau}_{\text{DI},1}$, where $\hat{\tau}_{\text{DI},j}$ for $j = 1, 2$ is derived by defining the following specific ψ_j function and solving the following estimating equations.

1. The ATE: Set $\psi_j(Z_i^{*(m)}, \tau_j) = I(G_i = j)\{R_{iT}Y_{\text{obs},iT} + (1 - R_{iT})Y_{\text{mis},iT}^{*(m)} - \tau_j\}$, and $\hat{\tau}_{\text{DI},j}$ is the solution to
$$\sum_{i=1}^{n} I(G_i = j)\left\{R_{iT}Y_{\text{obs},iT} + (1 - R_{iT})(M^{-1}\sum_{m=1}^{M} Y_{\text{mis},iT}^{*(m)}) - \tau_j\right\} = 0$$
2. The risk difference of the treatment: Set
$$\psi_j(Z_i^{*(m)}, \tau_j) = I(G_i = j)\left\{R_{iT}I(Y_{\text{obs},iT} \geq c) + (1 - R_{iT})I(Y_{\text{mis},iT}^{*(m)} \geq c) - \tau_j\right\}$$
and $\hat{\tau}_{\text{DI},j}$ is the solution to
$$\sum_{i=1}^{n} I(G_i = j)\left[R_{iT}I(Y_{\text{obs},iT} \geq c) + (1 - R_{iT})\left\{M^{-1}\sum_{m=1}^{M} I(Y_{\text{mis},iT}^{*(m)} \geq c)\right\} - \tau_j\right] = 0$$
3. Distributional information of the treatment, for example, the QTE for the qth quantile of responses when $\tau_{j,q}$ is the qth quantile for distribution of outcomes at the last time point: Set
$$\psi_j(Z_i^{*(m)}, \tau_{j,q}) = I(G_i = j)\left\{R_{iT}I(Y_{\text{obs},iT} \leq \tau_{j,q}) + (1 - R_{iT})I(Y_{\text{mis},iT}^{*(m)} \leq \tau_{j,q}) - q\right\}$$

and $\hat{\tau}_{\mathrm{DI},j}$ is the solution to

$$\sum_{i=1}^{n} I(G_i = j)\Big[R_{iT} I(Y_{\mathrm{obs},iT} \le \tau_{j,q}) + (1 - R_{iT})\Big\{M^{-1}\sum_{m=1}^{M} I(Y_{\mathrm{mis},iT}^{*(m)} \le \tau_{j,q})\Big\} - q\Big] = 0$$

Remark 1. (Discussion of the incorporation of the covariates in estimation) In sensitivity analyses, one may incorporate the covariate information to improve the efficiency of the treatment effect estimator.[30] For example, the ATE estimator derived from the sample average may not be the most efficient one; the one motivated by the analysis of covariance model (ANCOVA) is preferred in practice. We present an ANCOVA-motived DI estimator $\hat{\tau}_{\mathrm{DI},j}$ by defining the ψ_j function and its corresponding estimating equations for $j = 1, 2$ as follows:

Set the ψ_j function as

$$\psi_j(Z_i^{*(m)}, \tau_j, \gamma) = \begin{pmatrix} V_i\big\{R_{iT} Y_{\mathrm{obs},iT} + (1 - R_{iT}) Y_{\mathrm{mis},iT}^{*(m)} - V_i^{\mathrm{T}}\gamma\big\} \\ \tilde{X}_i^{\mathrm{T}} \quad I(j = 2)\tilde{X}_i^{\mathrm{T}}\gamma - \tau_j \end{pmatrix}$$

where $V_i = \left(\tilde{X}^{\mathrm{T}}, I(G_i = 2)\tilde{X}^{\mathrm{T}}\right)^{\mathrm{T}}$, and γ is the vector of the joint regression coefficients in the two treatment groups. $\hat{\tau}_{\mathrm{DI},j}$ is the solution to

$$\sum_{i=1}^{n}\begin{pmatrix} V_i\big\{R_{iT} Y_{\mathrm{obs},iT} + (1 - R_{iT})(M^{-1}\sum_{m=1}^{M} Y_{\mathrm{mis},iT}^{*(m)}) - V_i^{\mathrm{T}}\gamma\big\} \\ \tilde{X}_i^{\mathrm{T}} \quad I(j = 2)\tilde{X}_i^{\mathrm{T}}\gamma - \tau_j \end{pmatrix} = 0$$

Note that the ANCOVA-motivated estimator of the ATE replaces the treatment-specific covariate mean with the overall covariate mean to gain efficiency, while this information is not applicable to the risk difference and the QTE. Without this information, the estimating functions presented in Example 2 render the most efficient estimators of the ATE, the risk difference, and the QTE. One can use the augmented inverse propensity weighted (AIPW) type of estimators to incorporate additional information in the propensity score and outcome regression models[31]; however, the AIPW type of estimators does not improve the efficiency of the simple estimators (supported by unshown simulation studies).

To summarize, the DI procedure under specific sensitivity models is as follows.

Step 1. For each group, fit an MMRM from a population-averaged perspective for the observed data and obtain the model parameter estimator $\hat{\theta}$ by solving the estimating equation (4).

Step 2. Impute the missing values M times from the estimated imputation model $f(Z_{\mathrm{mis},i}^{*(m)} \mid Z_{\mathrm{obs},i}, \hat{\theta})$ based on prespecified imputation mechanisms such as Assumptions 1 to 3 for each group.

Step 3. Derive the DI estimator $\hat{\tau}_{\mathrm{DI},j}$ by solving the estimating equation (6) and get the treatment effect DI estimator $\hat{\tau}_{\mathrm{DI}}$.

The theoretical asymptotic properties of the DI estimator and the variance estimation procedure are given in Section 5.

Remark 2. (Computation complexity of DI and MI) Both DI and MI use the same imputation model to create multiple datasets in the imputation stage and use M^{-1} as the weight for each imputed dataset in the analysis stage. However, the approach to combining the imputed dataset and obtaining the treatment effect estimator after imputation is different. For MI, we conduct separate analyses for each imputed dataset and use Rubin's MI combining rule to get the inference; while for DI, the analysis is conducted jointly based on the entire imputed dataset, with the inference derived from the mean estimating equation (6) drawn on the idea of MC integration. In terms of the computation complexity after imputation, DI is more computationally efficient than MI for point estimation. For example, if linear models are fitted in the analysis stage, MI fits M linear models separately, with the total computational cost $O(Mnp^2 + Mp^3)$; DI fits one linear model for the pooled imputed dataset, with the total computational cost $O(Mnp^2 + p^3)$.[32]

Remark 3. (Connection with fractional imputation) DI is similar to parametric fractional imputation (FI),[33,34] where we pool the M imputed dataset and conduct the full-data analysis jointly by solving the estimating equations. Our proposed DI utilizes the distributional behavior of the missing components, by imputing them directly from the estimated conditional distribution given the observed data under some prespecified sensitivity analysis settings, thus avoiding the importance sampling required by FI to generate imputed data from the proposal distribution, and yields simplicity.

Remark 4. (Choice of the imputation size M) DI utilizes the idea of MC integration to approximate the conditional expectation in (3). Based on the MC approximation theory,[35] the MC error rate is $O(M^{-1/2})$ for any dimension. If the model is not

computationally intensive, larger M can be selected to further reduce the MC error. As shown in the simulation studies, the selection of the imputation size M is not sensitive to the inferences. We observe a decent performance of the DI estimator with a small imputation size M (e.g. $M = 5$).

5 Theoretical properties and variance estimation

5.1 Asymptotic properties of the DI estimator

We verify the consistency and asymptotic normality for the DI estimator. For simplicity, we consider the inference for one group here and omit the group subscript j. Extending to multiple groups is straightforward. Denote τ_0 as the true parameter such that $\mathbb{E}\{\psi(Z_i, \tau)\} = 0$. The comprehensive regularity conditions and technical proofs are given in Sections S1.1 and S1.2 in the Supplemental Material.

Theorem 1 Under the regularity conditions listed in Section S1.1 in the Supplemental Material, the DI estimator $\hat{\tau}_{\text{DI}} \mathbb{P} \tau_0$ as the imputation size $M \to \infty$ and sample size $n \to \infty$.

Theorem 2 Under the regularity conditions listed in Section S1.2 in the Supplemental Material, as the imputation size $M \to \infty$ and sample size $n \to \infty$,

$$\sqrt{n}(\hat{\tau}_{\text{DI}} - \tau_0) \xrightarrow{d} \mathcal{N}[0, A(\tau_0, \theta_0)^{-1} B(\tau_0, \theta_0) \{A(\tau_0, \theta_0)^{-1}\}^{\mathsf{T}}]$$

where

$$A(\tau_0, \theta_0) = \mathbb{E}\Big[\frac{\partial \psi(Z_i, \tau_0)}{\partial \tau} + \Big\{\frac{\partial \Gamma(\tau_0, \theta_0)}{\partial \tau}\Big\}\bar{s}_i(\theta_0)\Big];$$
$$B(\tau_0, \theta_0) = \mathbb{V}\big\{\psi_i^*(\tau_0, \theta_0) + \Gamma(\tau_0, \theta_0)^{\mathsf{T}} \bar{s}_i(\theta_0)\big\}$$

Here $\psi_i^*(\tau, \theta) = M^{-1} \sum_{m=1}^{M} \psi(Z_i^{*(m)}, \tau)$, $\bar{s}_i(\theta) = \mathbb{E}\{s(Z_i, \theta) \mid Z_{\text{obs},i}\}$, $\Gamma(\tau, \theta_0) = \mathbf{I}_{\text{obs}}(\theta_0)^{-1} \mathbf{I}_{\psi,\text{mis}}(\tau, \theta_0)$, where $\mathbf{I}_{\text{obs}}(\theta) = \mathbb{E}\{-\partial \bar{s}_i(\theta)/\partial \theta\}$ and $\mathbf{I}_{\psi,\text{mis}}(\tau, \theta) = \mathbb{E}[\{s(Z_i, \theta) - \bar{s}_i(\theta)\}\psi(Z_i, \tau)]$.

5.2 Variance estimation

From the result of asymptotic normality in Theorem 2, one can plug the model parameter estimator $\hat{\theta}$ and the DI estimator $\hat{\tau}_{\text{DI}}$ in the asymptotic variance formula to obtain a consistent variance estimator $\hat{\mathbb{V}}_1(\hat{\tau}_{\text{DI}})$, which is given in Section S1.3 in the Supplemental Material. However, $\hat{\mathbb{V}}_1(\hat{\tau}_{\text{DI}})$ involves the analytical form of the estimated score function $s(Z_i^{*(m)}, \hat{\theta})$ that is difficult to compute in longitudinal settings. The replication-based variance estimation is preferred for its simplicity, and it is commonly obtained by nonparametric bootstrap. But it requires computational efforts and the re-imputation of the missing components on the refitted imputation model, especially in a large-scale clinical trial with numerous participants. We propose weighted bootstrap to obtain a consistent variance estimator without having to re-impute the missing values per bootstrap iteration.

The weighted bootstrap procedure has parallel steps as the DI procedure. However, cautions should be taken when deriving the replicated DI estimator. To preserve the imputation model in DI, we draw on the idea of importance sampling[35] to incorporate the variability of the current replicated model parameter estimator $\hat{\theta}^{(b)}$ and the target parameter estimator $\hat{\tau}_{\text{DI}}^{(b)}$ in each bootstrap iteration $b = 1, \ldots, B$, where B is the total number of bootstrap replicates. A standard recommendation for the total number of bootstrap replicates to get the variance estimation is $B = 100$.[36] In this way, one can approximate the conditional expectation $\mathbb{E}\{\psi(Z_i, \tau) | Z_{\text{obs},i}, \hat{\theta}^{(b)}\}$ by a weighted sum as

$$\mathbb{E}\big\{\psi(Z_i, \tau) \mid Z_{\text{obs},i}, \hat{\theta}^{(b)}\big\} \approx \sum_{m=1}^{M} w_i^{(m)}(\hat{\theta}^{(b)}) \psi(Z_i^{*(m)}, \tau)$$

where the importance weights are computed as

$$w_i^{(m)}(\hat{\theta}^{(b)}) \propto \frac{f(Z_i^{*(m)} \mid Z_{\text{obs},i}, \hat{\theta}^{(b)})}{f(Z_i^{*(m)} \mid Z_{\text{obs},i}, \hat{\theta})} \tag{7}$$

with the constraint $\sum_{m=1}^{M} w_i^{(m)}(\hat{\theta}^{(b)}) = 1$ for all i.

To summarize, we conduct the weighted bootstrap procedure in each iteration as follows.

Step 1. Generate the i.i.d. bootstrap weights $u_i^{(b)}$ such that $\mathbb{E}(u_i^{(b)}) = 1$ and $\mathbb{V}(u_i^{(b)}) = 1$, with $u_i^{(b)} \geq 0$ for each individual. Obtain the model parameter estimator $\hat{\theta}^{(b)}$ by solving the estimating equations

$$\sum_{i=1}^{n} u_i^{(b)} \mathbb{E}\{s(Z_i, \theta) \mid Z_{\text{obs},i}\} = 0 \tag{8}$$

Step 2. Update the importance weights $w_i^{(m)}(\hat{\theta}^{(b)})$ such that it satisfies (7) with a constraint $\sum_{m=1}^{M} w_i^{(m)}(\hat{\theta}^{(b)}) = 1$, under the prespecified imputation settings such as Assumptions 1 to 3.

Step 3. Obtain the DI estimator $\hat{\tau}_{\text{DI}}^{(b)}$ by solving the estimating equations

$$\sum_{i=1}^{n} \sum_{m=1}^{M} u_i^{(b)} w_i^{(m)}(\hat{\theta}^{(b)}) \psi(Z_i^{*(m)}, \tau) = 0 \tag{9}$$

Repeat Steps 1 to 3 B times, and get the replication variance estimator of the DI estimator as

$$\hat{\mathbb{V}}_2(\hat{\tau}_{\text{DI}}) = \frac{1}{B-1} \sum_{b=1}^{B} (\hat{\tau}_{\text{DI}}^{(b)} - \hat{\tau}_{\text{DI}})^2$$

Remark 5. (Choice of the weight distribution) There are many candidate distributions to generate the bootstrap weights $u_i^{(b)}$. For example, one may try an exponential distribution with the rate parameter 1 denoted as Exp(1), or a discrete distribution, such as Poisson with parameter 1. The choice of the generated distribution is not sensitive to the variance estimation. We adopt Exp(1) in simulation studies.

Theorem 3 shows the asymptotic validity of the above weighted bootstrap method, with the proof in Section S1.4 in the Supplemental Material.

Theorem 3. Under regularity conditions listed in Sections S1.1 and S1.2 in the Supplemental Material, with the bootstrap weights $u_1^{(b)}, \ldots, u_n^{(b)}$ i.i.d. satisfying $\mathbb{E}(u_i^{(b)}) = 1$, $\mathbb{V}(u_i^{(b)}) = 1$ with $u_i^{(b)} \geq 0$, $\hat{\mathbb{V}}_2(\hat{\tau}_{\text{DI}})$ is a consistent estimator of $\mathbb{V}(\hat{\tau}_{\text{DI}})$.

6 Simulation study

We conduct simulation studies to assess the finite-sample validity of our proposed DI framework for sensitivity analyses in longitudinal clinical trials. Consider a clinical trial with two groups and five visits. The baseline covariates X are generated from the standard normal distribution with $p = 3$ dimensions. For the longitudinal responses Y_i of the ith individual, we generate them independently from a multivariate normal distribution as the full-data model (1), where for the control and treatment group, the group-specific coefficients $\beta_{j1}, \ldots, \beta_{j5}$ and covariance matrices $\Sigma^{(j)}$ for $j = 1, 2$ are presented in Section S2 in the Supplemental Material.

Consider the missing mechanism as MAR with a monotone missingness pattern. Specifically, assume all the baseline responses are observed, that is, $R_{i1} = 1$. For visit $k > 1$, if $R_{ik-1} = 0$, then $R_{il} = 0$ for $l = k, \ldots, T$; otherwise let $R_{ik} \sim \text{Bernoulli}(\pi_{ik})$. Model the observed probability π_{ik} at visit $k > 1$ as a function of the observed information as $\text{logit}(\pi_{ik}|G_i = j) = \phi_{1j} + \phi_{2j} Y_{ik-1}$, where ϕ_{1j}, ϕ_{2j} are tuning parameters for the observed probabilities. We set $\phi_{11} = -3.2$, $\phi_{12} = -4.0$, $\phi_{21} = \phi_{22} = 0.2$ to get the observed probabilities as 0.7865 and 0.7938 for the control and treatment groups, respectively.

Different types of treatment effect estimands including the ATE, the risk difference, and the QTE are assessed. When the primary interest is the ATE, we use the ATE estimator motivated by ANCOVA since it shows an increase in efficiency compared to the one obtained by directly taking the sample average. When the risk difference is of interest, we set a

threshold as $c = 4.5$ and are interested in $\tau = P(Y_{iT} \geq 4.5|G_i = 2) - P(Y_{iT} \geq 4.5|G_i = 1)$. When the QTE is of interest, we set $q = 0.5$ to obtain the behavior of the median. We focus on the sensitivity analysis under J2R to describe the deviation from MAR, which matches the setting in our motivating example. For illustration, we only present the result for the ATE under J2R. Simulation results for sensitivity analyses under other sensitivity models such as RTB and washout imputation, along with other treatment effect estimands under J2R are given in Section S2 in the Supplemental Material. We select the number of bootstrap replicates $B = 100$. Consider the sample size N for each group to be the same value ranging from $\{100, 500, 1000\}$ in each group, and the imputation size M ranging from $\{5, 10, 100\}$. In addition, we select the distribution Exp(1) to generate the bootstrap weights.

We compared our proposed DI with MI in the simulation study. Rubin's method and weighted bootstrap are applied to the MI and DI estimators to get the variance estimation, respectively, under 1000 MC simulations. The estimators are assessed using the point estimate (Point est), the MC variance denoted as true variance (True var), the variance estimate (Var est), the relative bias of the variance estimate computed by $[\mathbb{E}\{\hat{\mathbb{V}}(\hat{\tau})\} - \mathbb{V}(\hat{\tau})]/\mathbb{V}(\hat{\tau})$, the coverage rate of 95% confidence interval (CI), and mean CI length. We choose the 95% Wald CI estimated by $(\hat{\tau} - 1.96\hat{\mathbb{V}}^{1/2}(\hat{\tau}), \hat{\tau} + 1.96\hat{\mathbb{V}}^{1/2}(\hat{\tau}))$.

Table 1 shows the simulation result of the ATE estimator. The point estimates from both DI and MI are closer to the true value as the sample size increases, and their MC variances are getting smaller. It indicates that the MI and DI estimators are consistent and have comparable performances. The efficiency of the estimator increases as the imputation size M grows. For variance estimation, Rubin's method ends up overestimating the true variance as expected, with a larger relative bias and a conservative coverage rate. The variance estimate using weighted bootstrap in DI is close to the true variance, with a well-controlled relative bias and a coverage rate close to the empirical value. For other types of estimands, the variance estimate of the QTE in MI and DI overestimates the true variance when the sample size is small. But as the sample size grows, the results from DI are much better and proximal to the true value. Therefore, relatively large sample size is suggested when estimating the QTE based on the simulation results.

Each type of treatment effect estimator based on DI and MI has comparable performance in terms of the point estimation under each prespecified sensitivity model. The variance estimation using weighted bootstrap in DI outperforms Rubin's MI combining method in all cases with much tolerable relative biases and better coverage probabilities. The same interpretations of the results apply to other settings as shown in Section S2 in the Supplemental Material.

7 Return to the motivating example

We apply our proposed DI framework to the motivating example in Section 2 to uncover the treatment effect in sensitivity analyses. Apart from comparing the performance between MI with Rubin's rule and DI with the proposed weighted bootstrap for the ATE and the risk difference, we also explore the QTE defined by the quantile difference between the relative change of the HAMD-17 score in the sensitivity analysis under J2R. The three treatment effect estimands are estimated through the estimating equations in Example 2 after imputation. For the QTE, we do not limit it to one specific quantile; instead, we present the estimated cumulative distribution function (CDF) of the relative change from baseline in each group. In the implementation of MI and DI, we set the imputation size as $M = 100$ and the number of bootstrap replicates as $B = 100$.

Table 1. Simulation results under J2R of the ATE estimator. Here the true value $\tau = 1.5400$.

		Point est		True var		Var est		Relative bias		Coverage rate		Mean CI length	
		($\times 10^{-2}$)		($\times 10^{-2}$)		($\times 10^{-2}$)		(%)		(%)		($\times 10^{-2}$)	
N	M	MI	DI	MI	DI	MI	DI	MI	DI	MI	DI	MI	DI
100	5	150.84	150.36	14.86	14.60	20.92	14.70	40.85	0.74	97.90	94.90	178.76	149.61
	10	150.95	150.48	14.27	14.25	20.63	14.43	44.59	1.25	97.90	94.90	177.64	148.19
	100	150.74	150.76	14.00	13.91	20.37	14.21	45.44	2.17	97.80	95.10	176.61	147.05
500	5	153.52	153.13	3.06	3.07	4.08	3.03	33.44	−1.25	98.00	94.50	79.09	67.99
	10	153.42	153.45	3.02	3.02	4.01	2.97	33.09	−1.43	97.80	94.50	78.48	67.38
	100	153.39	153.43	2.98	2.99	3.97	2.92	33.17	−2.27	98.20	94.10	78.07	66.75
1000	5	154.27	154.06	1.46	1.47	2.03	1.52	39.23	3.48	97.60	94.90	55.84	48.14
	10	154.09	154.06	1.45	1.44	2.00	1.49	38.63	3.54	97.70	94.80	55.45	47.70
	100	154.12	154.11	1.43	1.42	1.98	1.46	38.55	2.99	97.90	94.20	55.18	47.26

J2R: jump-to-reference; MI: multiple imputation; DI: distributional imputation; ATE: average treatment effect.

Table 2. Analysis of the HAMD-17 data in terms of the ATE.

	Point estimation		Standard error		*P*-value	
Setting	MI (95% CI)	DI (95% CI)	MI	DI	MI	DI
MAR	−2.42 (−4.49, −0.35)	−2.30 (−4.47, −0.13)	1.06	1.11	0.022	0.038
J2R	−1.81 (−3.91, 0.29)	−1.68 (−3.28, −0.08)	1.07	0.82	0.091	0.039
RTB	−1.23 (−3.39, 0.93)	−1.25 (−3.13, 0.63)	1.10	0.96	0.266	0.192
Washout	−0.71 (−2.83, 1.41)	−0.75 (−2.79, 1.39)	1.08	1.04	0.511	0.475

HAMD-17: Hamilton Depression Rating Scale for 17 items; MAR: missing at random; RTB: return-to-baseline; J2R: jump-to-reference; MI: multiple imputation; DI: distributional imputation; ATE: average treatment effect.

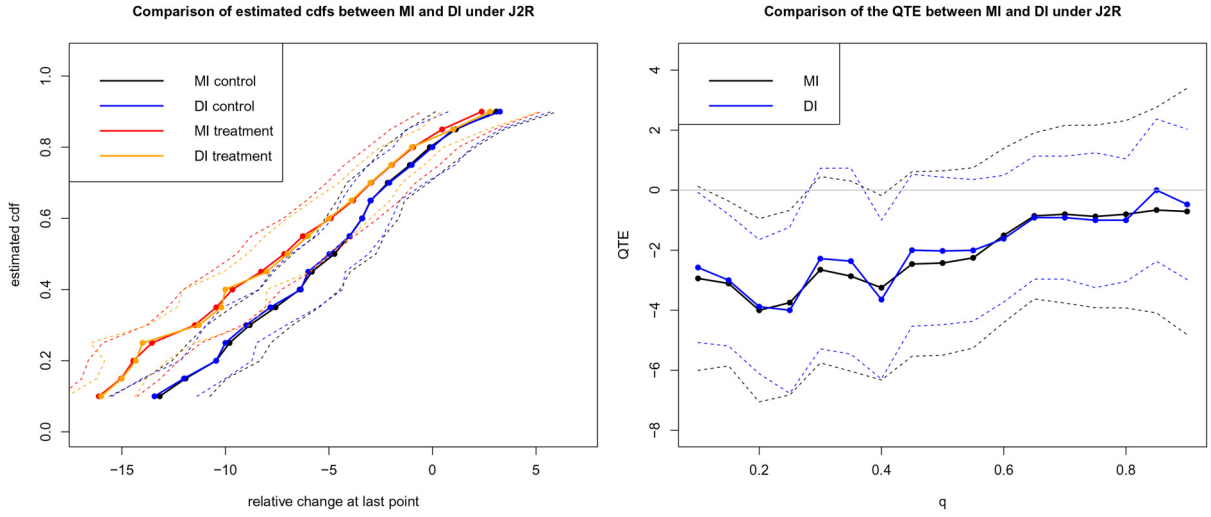

Figure 2. The estimated CDF and QTE of the relative change from baseline at the last visit via MI and DI in the sensitivity analysis under J2R, accompanied by the point-wise 95% CI in dashed lines. CDF: cumulative distribution function; QTE: quantile treatment effect; MI: multiple imputation; DI: distributional imputation.

Table 2 and Table S9 in the Supplemental Material present the analysis results under MI and DI. The primary analysis in the "MAR" rows in the tables indicates a parallel performance of MI and DI in terms of the point and variance estimates, validating the consistency of both variance estimators under MAR. While MI with Rubin's variance estimator alters the study conclusion under J2R at the significance level $\alpha = 0.05$, DI with the weighted bootstrap procedure preserves a significant treatment effect by producing smaller standard errors and narrower CIs. Applying DI with weighted bootstrap resolves the suspicion of the effectiveness of the experimental drug, as it guarantees a consistent variance estimate of the treatment effect.

To evaluate the distributional behavior of the data, we estimate the CDF of the relative change of the HAMD-17 score at the last time point in each group and the QTE as a function of q denoted as the quantile percentage under J2R in Figure 2. Similar to the results of the ATE and the risk difference, the estimated CDF obtained from DI has a comparable shape to the one obtained from MI, while the curve from DI has a narrower 95% confidence region in the sensitivity analysis compared to MI. A statistically significant treatment effect is detected for patients in the lower quantiles of the HAMD-17 score.

Our general framework captures a comprehensive evaluation of the treatment effect. With the use of DI, the experimental drug reveals its significant benefit for curing depression in both the primary analysis and the sensitivity analysis under J2R. If other sensitivity models such as RTB and washout imputation are assumed in the trial protocol, we also provide the corresponding results in Table 2 and in Section S3.2 in the Supplemental Material, to illustrate a wide application of the proposed framework. Under each sensitivity model, DI produces smaller standard errors and CIs compared to MI. However, one should notice that the study conclusion regarding the treatment effect changes with the prespecified sensitivity models. Among those sensitivity analyses, only the result under J2R using DI captures the same statistical significance of the ATE; the result under washout imputation fails to show an improvement of the treatment with respect to

the risk difference. It suggests a potential impact of different missingness assumptions on the treatment effect of an experimental drug. Statisticians should carefully interpret the primary and sensitivity analysis results with investigators based on the missingness assumptions and additional clinical knowledge of the drug.

8 Conclusion

In this paper, we propose DI using the idea of MC integration and establish a unified framework for sensitivity analyses in longitudinal clinical trials to assess the impact of the MAR assumption in the primary analysis. Our framework is flexible to accommodate various sensitivity models and treatment effect estimands. We apply the proposed DI approach with weighted bootstrap to an antidepressant longitudinal clinical study and detect a statistically significant treatment effect in both the primary and sensitivity analyses under J2R for the ATE, the risk difference, and the QTE, overcoming the inefficiency and overestimation issue from MI with Rubin's rule. The study result of the experimental drug uncovers a significant improvement for curing depression. The DI framework has a solid theoretical guarantee, with the avoidance of re-imputation of the missing data in the variance estimation via weighted bootstrap.

While we present DI under a monotone missingness pattern, DI is applicable to handle intermittent missing values as long as the imputation model of the missing values given the observed values is tractable. If direct sampling from the target imputation model is difficult, one can resort to alternative sampling strategies such as importance sampling, Metropolis-Hastings, etc. We leave this topic as a future research direction.

Although we present the framework for sensitivity analyses using continuous longitudinal outcomes, it is possible to extend the framework to the cases of categorical or survival outcomes under some prespecified imputation mechanisms. For example, Tang[37] modifies the control-based imputation model for binary and ordinal responses based on the generalized linear mixed model; Yang et al.[38] adopt the δ-adjusted and control-based imputation models for survival outcomes in sensitivity analyses. With motivated treatment effect estimands and suitable prespecified sensitivity assumptions, our framework can handle sensitivity analyses for different types of outcomes in clinical trials. Guan and Yang[39] establish a unified framework of MI via wild bootstrap for causal inference in observational studies; extending the proposed DI inference to this context is straightforward.

The established general framework for sensitivity analyses in longitudinal clinical trials involves the normality assumption in both the imputation and analysis stages. This parametric configuration originates from the MI-based approaches in sensitivity analyses.[3,25] When the data violates the normality, using the normal distribution for imputation is still robust if the estimator of interest is asymptotically normal.[40,11] In the future, we will develop DI under semiparametric and nonparametric models as more flexible settings in sensitivity analyses.

Acknowledgements

Yang is partially supported by the NSF grant DMS 1811245, NIA grant 1R01AG066883, and NIEHS grant 1R01ES031651.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iDs

Shu Yang https://orcid.org/0000-0001-7703-707X
Yilong Zhang https://orcid.org/0000-0002-0668-4429

Supplemental material

The supplemental material include the setup and proof for the theorems, additional simulation and real data application results, and the codes for the DI implementation.

References

1. ICH. E9(R1) statistical principles for clinical trials: addendum: estimands and sensitivity analysis in clinical trials. *FDA Guidance Documents* 2021.
2. Rubin DB. Inference and missing data. *Biometrika* 1976; **63**: 581–592.

3. Carpenter JR, Roger JH and Kenward MG. Analysis of longitudinal trials with protocol deviation: a framework for relevant, accessible assumptions, and inference via multiple imputation. *J Biopharm Stat* 2013; **23**: 1352–1371.
4. Little RJ, D'Agostino R, Cohen ML, et al. The prevention and treatment of missing data in clinical trials. *N Engl J Med* 2012; **367**: 1355–1360.
5. Scharfstein DO, Rotnitzky A and Robins JM. Adjusting for nonignorable drop-out using semiparametric nonresponse models. *J Am Stat Assoc* 1999; **94**: 1096–1120.
6. Robins JM, Rotnitzky A and Scharfstein DO. Sensitivity analysis for selection bias and unmeasured confounding in missing data and causal inference models. In *Statistical models in epidemiology, the environment, and clinical trials*. Springer, 2000. pp. 1–94.
7. Yin P and Shi JQ. Simulation-based sensitivity analysis for non-ignorably missing data. *Stat Methods Med Res* 2019; **28**: 289–308.
8. Steenland K and Greenland S. Monte Carlo sensitivity analysis and Bayesian analysis of smoking as an unmeasured confounder in a study of silica and lung cancer. *Am J Epidemiol* 2004; **160**: 384–392.
9. Little RJ. Pattern-mixture models for multivariate incomplete data. *J Am Stat Assoc* 1993; **88**: 125–134.
10. Little RJ and Wang Y. Pattern-mixture models for multivariate incomplete data with covariates. *Biometrics* 1996; **52**: 98–111.
11. Cro S, Morris TP, Kenward MG, et al. Sensitivity analysis for clinical trials with missing continuous outcome data using controlled multiple imputation: a practical guide. *Stat Med* 2020; **39**: 2815–2842.
12. US Food and Drug Administration. Statistical review and evaluation of tresiba and ryzodeg 70/30, 2016. https://www.fda.gov/media/102782/download.
13. Mehrotra DV, Liu F and Permutt T. Missing data in clinical trials: control-based mean imputation and sensitivity analysis. *Pharm Stat* 2017; **16**: 378–392.
14. Zhang Y, Golm G and Liu G. A likelihood-based approach for the analysis of longitudinal clinical trials with return-to-baseline imputation. *Stat Biosci* 2020; **12**: 23–36.
15. Rubin DB. *Multiple imputation for nonresponse in surveys*. 81, New York: John Wiley & Sons, 2004.
16. Meng XL. Multiple-imputation inferences with uncongenial sources of input. *Stat Sci* 1994; **10**: 538–558.
17. Robins JM and Wang N. Inference for imputation estimators. *Biometrika* 2000; **87**: 113–124.
18. Yang S and Kim JK. A note on multiple imputation for method of moments estimation. *Biometrika* 2016; **103**: 244–251.
19. Liu GF and Pang L. On analysis of longitudinal clinical trials with missing data using reference-based imputation. *J Biopharm Stat* 2016; **26**: 924–936.
20. Lepage GP. A new algorithm for adaptive multidimensional integration. *J Comput Phys* 1978; **27**: 192–203.
21. Mallinckrodt C, Roger J, Chuang-Stein C, et al. Recent developments in the prevention and treatment of missing data. *Ther Innov Regul Sci* 2014; **48**: 68–80.
22. Roussel R, Duran-García S, Zhang Y, et al. Double-blind, randomized clinical trial comparing the efficacy and safety of continuing or discontinuing the dipeptidyl peptidase-4 inhibitor sitagliptin when initiating insulin glargine therapy in patients with type 2 diabetes: The composit-i study. *Diabetes Obes Metab* 2019; **21**: 781–790.
23. Yang S and Zhang Y. Multiply robust matching estimators of average and quantile treatment effects. *Scand J Stat* 2020; **49**: 1–31.
24. Mallinckrodt C, Molenberghs G, Lipkovich I et al. *Estimands, estimators and sensitivity analysis in clinical trials*. London: Chapman and Hall/CRC, 2019.
25. Zhang Y, Zimmer Z, Xu L, et al. Missing data imputation with baseline information in longitudinal clinical trials. *Stat Biopharm Res* 2022; **14**: 242–248.
26. US Food and Drug Administration. Statistical review and evaluation of praluent (alirocumab), 2018. https://www.accessdata.fda.gov/drugsatfda˙docs/nda/2018/125559Orig1s014StatR.pdf.
27. Mallinckrodt C, Bell J, Liu G, et al. Aligning estimators with estimands in clinical trials: putting the ich e9 (r1) guidelines into practice. *Ther Innov Regul Sci* 2020; **54**: 353–364.
28. Hughes RA, Sterne J and Tilling K. Comparison of imputation variance estimators. *Stat Methods Med Res* 2016; **25**: 2541–2557.
29. Louis TA. Finding the observed information matrix when using the em algorithm. *J R Stat Soc: Ser B (Methodological)* 1982; **44**: 226–233.
30. Tsiatis A. *Semiparametric theory and missing data*. New York: Springer Science & Business Media, 2007.
31. Zhang Z, Chen Z, Troendle JF, et al. Causal inference on quantiles with an obstetric application. *Biometrics* 2012; **68**: 697–706.
32. Friedman J, Hastie T, Tibshirani R, et al. *The elements of statistical learning*, volume 1. Springer series in statistics, New York, 2001.
33. Kim JK. Parametric fractional imputation for missing data analysis. *Biometrika* 2011; **98**: 119–132.
34. Yang S and Kim JK. Fractional imputation in survey sampling: a comparative review. *Stat Sci* 2016; **31**: 415–432.
35. Geweke J. Bayesian inference in econometric models using monte carlo integration. *Econometrica: J Economet Soc* 1989; **1**: 1317–1339.
36. Boos DD and Stefanski LA. *Essential statistical inference: theory and methods*. 120, New York: Springer Science & Business Media, 2013.
37. Tang Y. Controlled pattern imputation for sensitivity analysis of longitudinal binary and ordinal outcomes with nonignorable dropout. *Stat Med* 2018; **37**: 1467–1481.
38. Yang S, Zhang Y, Liu GF, et al. SMIM: a unified framework of survival sensitivity analysis using multiple imputation and martingale. *Biometrics* 2021; **77**: 1–11. DOI: 10.1111/biom.13555.
39. Guan Q and Yang S. A unified framework for causal inference with multiple imputation using martingale. *arXiv preprint arXiv:191104663* 2019.
40. Schafer JL. *Analysis of incomplete multivariate data*. London: CRC Press, 1997.