How to relate potential outcomes: Estimating individual treatment effects under a given specified partial correlation

Mingyang Cai, Stef van Buuren and Gerko Vink

Abstract

In most medical research, the average treatment effect is used to evaluate a treatment's performance. However, precision medicine requires knowledge of individual treatment effects: What is the difference between a unit's measurement under treatment and control conditions? In most treatment effect studies, such answers are not possible, because the outcomes under both experimental conditions are not jointly observed. This makes the problem of causal inference a missing data problem. We propose to solve this problem by imputing the individual potential outcomes under a specified partial correlation (SPC), thereby allowing for heterogeneous treatment effects. We demonstrate in simulation that our proposed methodology yields valid inference for the marginal distribution of potential outcomes. We highlight that the posterior distribution of individual treatment effects varies with different specified partial correlations. This property can be used to

study the sensitivity of optimal treatment outcomes under different correlation specifications. In a practical example on HIV-1 treatment data, we demonstrate that the proposed methodology generalises to real-world data. Imputing under the SPC therefore opens up a wealth of possibilities for studying heterogeneous treatment effects on incomplete data and the further adaptation of individual treatment effects.

Keywords: Multiple imputation, joint modeling imputation, iterative imputation, multivariate data analysis

1 Introduction

Heterogeneity of treatment effects across individuals is a significant complication in precision treatment assignment to different persons. The difficulty of evaluating individual treatment effects (ITE) from the observed data is that only one of the potential outcomes is observed for each individual (Rubin, 1974; Hernan & Robins, 2010). This fundamental problem of causal inference implies that causal inference is essentially a missing data problem (Rubin, 2005; Peng Ding, 2018). Simply ignoring the missingness in the potential outcomes would only allow for average or homogeneous treatment effects. To allow for the estimation of unobserved heterogeneous treatment effects, we need to solve for the individual missing potential outcomes through multiple imputation.

Multiple imputation is a popular approach for analysing incomplete datasets but is not yet widely used in causal inference. In multiple imputation the missingness is solved before the data is analysed as if it were completely

observed. Imputations for missing values are therein drawn from the corresponding posterior predictive distributions in parallel, resulting in multiple imputed datasets. Then, the statistical inference is obtained for each imputed dataset separately by complete-data analyses. Finally, the multiple analyses are aggregated into a single inference using Rubin's rules (Rubin, 1987, pp. 76), which account for within, and across imputation uncertainty. Because multiple imputation imputes the individual potential outcomes, we can evaluate both the difference between outcomes and the individual treatment effects.

Studying individual treatment effects receives increasing attention. For example, Lamont et al. (2018), and Westreich et al. (2015) discussed the performance of multiple imputation for potential outcomes. Lamont et al. (2018) applied multiple imputation to evaluate the effectiveness of various programs designed to prevent depression among sampled women and provide program recommendations for women out of the sample. However, Lamont et al. (2018) and Westreich et al. (2015) fitted separate imputation models for potential outcomes based on observed covariates, thereby implicitly assuming conditional independence between potential outcomes. This conditional independence assumption is not always valid and cannot be verified from the observed data (Rässler, 2012). Imbens and Rubin (2015) and Gadbury et al. (2001) studied the sensitivity of the average treatment effect estimates under violations of the assumption of conditional independence between potential outcomes. They found that distributions of average causal effect under various partial correlations are different. An alternative imputation strategy fits a fully conditional model for the incomplete outcome. However, Van Buuren

3

(2018) demonstrated that without the specification of the partial correlation, the derived imputations for such models are unstable and can be implausible.

Specification of the partial correlation in applications of multiple imputation for potential outcomes has received little attention to date. We know that the partial correlation could be an arbitrary value between -1 and 1 in each imputed dataset. However, the imputations become poor when the partial correlation in the imputed dataset is negative (Van Buuren, 2018, section 8.4.1). We therefore assume that this correlation is non-negative. Smink (2016) proposed a data augmentation approach, where rows are added to the data that hold prior information for the partial correlation. This procedure is also outlined in Van Buuren (2018, section 8.4.2). In Smink's scenario, the imputations are guided by the specified correlation in the augmented cases, but the data does not hold any covariates. One could imagine that augmenting the data with a joint prior set becomes increasingly challenging when the number of covariates increases.

We propose a new hybrid imputation approach for imputing potential outcomes under a given partial correlation that allows for the collection of incomplete covariates. The procedure is hybrid in the sense that it combines properties from joint modeling imputation (the potential outcomes form a joint and are imputed as such) and fully conditional specification, wherein the covariates are imputed on a fully conditional variable-by-variable basis. In this manuscript we first outline the role of the partial correlation in causal inference, then give a brief overview of multiple imputation and introduce our new hybrid imputation approach. We evaluate the validity of the methodology in simulation and demonstrate the real-world applicability

4

on a clinical trial aimed to evaluate the individual treatment effects of two different therapies on slowing the progression of HIV disease.

2 The role of partial correlation

2.1 Notation

Let $Y(j), j = 1, \ldots, p$ denote one of p incomplete variables and $Y(-j) = (Y(1), \ldots, Y(j-1), Y(j+1), \ldots, Y(p))$ denote the collection of the $p-1$ variables in Y except $Y(j)$. In this paper, $Y(j)$ ususally represents the potential outcomes. Let $X = (X(1), \ldots, X(k))$ be a set of k completely observed variables. Let $\rho(Y(0), Y(1))$ be the correlation between two potential outcomes and $\rho_{Y(0), Y(1) \mid X}$ be the partial correlation between two potential outcomes.

2.2 Setup

We focus on the case of a binary treatment W_i and a continuous outcome Y_i and assume the data come from a random sample of individuals, indexed by $i \in 1, \ldots, N$. Each individual i has a nonzero probability to be assigned to both treatments, with $W_i = 1$ for the active treatment and $W_i = 0$ for the control treatment. The number of units under treatment and control are N_1 and N_0 respectively. We assume that the treatment assignment mechanism is unconfounded by the unobserved outcomes Y_{mis}, i.e., an ignorable assignment mechanism. We also assume that the potential outcomes for any individual are independent of the treatments assigned to others, which

is known as the *stable unit treatment value assumption* (Imbens & Rubin, 2015). Here the ignorable or unconfounded assignment mechanism implies that $P(W|Y(0), Y(1), X) = P(W|Y_{obs}, X)$, where X are observed covariates not influenced by treatment assignment. The individual treatment effect is defined as $\tau_i = Y_i(1) - Y_i(0)$. We assume a joint distribution for the potential outcomes Y_1 and Y_0 and that the correlation between the potential outcomes $\rho(Y_1, Y_0)$ can be quantified by one or more parameters. The imputation models for missing outcomes $Y(1)$ and $Y(0)$ are:

$$\dot{Y}(1) \sim P(Y^{mis}(1)|Y^{obs}(1), Y(0), X, \dot{\phi}_1) \tag{1}$$

$$\dot{Y}(0) \sim P(Y^{mis}(0)|Y^{obs}(0), Y(1), X, \dot{\phi}_0), \tag{2}$$

where parameters of the imputation model $\dot{\phi}_1$ and $\dot{\phi}_0$ are draws from their respective posterior distribution.

2.3 Partial correlation between potential outcomes

The necessity of specifying the partial correlation in the process of multiple imputation has been discussed hereinbefore. This section would further illustrate the causality meaning of the partial correlation. We decompose the individual treatment effect $\tau_i \in T, i = 1, \ldots, N$ via

$$\tau_i = Y_i(1) - Y_i(0) = X_i^{\mathsf{T}}\beta + \varepsilon_i, \tag{3}$$

where X_i are observed pre-treatment covariates for individual i. Under random treatment assignment, the regression weight β could be the ordinary

least-square (OLS) estimation of τ_i on X_i. The quantity $X_i^\mathsf{T}\beta$ is known as systematic treatment effect variation and the residual ε_i is the idiosyncratic treatment effect variation not explained by X_i (Ding et al., 2019; Heckman et al., 1997; Djebbari & Smith, 2008). The idiosyncratic variation accounts for treatment effect variation not attributable to differences in observed covariates. Based on the formula of OLS estimation, the coefficient

$$
\begin{aligned}
\beta &= (X^\mathsf{T}X)^{-1}X^\mathsf{T}T \\
&= (X^\mathsf{T}X)^{-1}X^\mathsf{T}Y(1) - (X^\mathsf{T}X)^{-1}X^\mathsf{T}Y(0) \\
&= \beta_1 - \beta_0,
\end{aligned}
\tag{4}
$$

where β_1 and β_0 are the corresponding regression weights of the potential outcomes Y_1 and Y_0 on the observed covariates X. Similarly, the idiosyncratic treatment effect variation ε_i

$$
\begin{aligned}
\varepsilon_i &= \tau_i - X_i^\mathsf{T}\beta \\
&= (Y_i(1) - X_i^\mathsf{T}\beta_1) - (Y_i(0) - X_i^\mathsf{T}\beta_0) \\
&= \varepsilon_i(1) - \varepsilon_i(0),
\end{aligned}
\tag{5}
$$

where $\varepsilon(1)$ and $\varepsilon(0)$ are the residuals from the regression of the potential outcomes Y_1 and Y_0 on the observed covariates X. Applying the theory of variance decomposition for linear regression, we could decompose the variance of individual treatment effect into two components:

$$
\mathrm{Var}(\tau_i) = \mathrm{Var}(X_i^\mathsf{T}\beta) + \mathrm{Var}(\varepsilon_i).
\tag{6}
$$

Based on the idiosyncratic treatment effect variation formula (5), the idiosyncratic components of individual treatment variance becomes

$$
\begin{aligned}
\mathrm{Var}(\varepsilon_i) \;&= \mathrm{Var}[\varepsilon_i(1) - \varepsilon_i(0)] \\
&= \mathrm{Var}[\varepsilon_i(1)] + \mathrm{Var}[\varepsilon_i(0)] - 2\mathrm{Cov}[\varepsilon_i(1), \varepsilon_i(0)],
\end{aligned}
\tag{7}
$$

which demonstrate that partial correlation $\rho_{Y(0)Y(1)\mid X}$ between the potential outcomes does impact the idiosyncratic variation, which is unidentifiable from the observed data.

Although the relation between potential outcomes cannot be determined solely from the observed data, there are still some approaches to identifying the partial correlation, such as the model-based approach, the experiment-based approach, and sensitivity analysis. The model-based approach explicitly models the relationship between the idiosyncratic variation and the assignment mechanisms such that the partial correlation would be close or equal to zero (Heckman, 2005). Economists usually deal with ex-post causal inference. The agents select the treatment according to their ex-ante evaluation. In this case, economists could infer the information only available to agents from the assignment mechanism. For instance, Heckman (2010) investigated the causal effect of educational decisions on the labour market and health outcomes. He modelled latent cognitive and social-emotional endowments and included these latent variables into the outcome equations. As indicated earlier, the partial correlation between potential outcomes is the only unknown parameter in the formula of idiosyncratic variation. Therefore, the model for idiosyncratic variation could be tailored to the model for

the partial correlation between potential outcomes. Generally, the model-based approach attempts to figure out latent variables affecting the partial correlation between potential outcomes.

The experiment-based approach intends to design sophisticated experiments to collect additional data where analysts could evaluate the partial correlation between potential outcomes. For instance, the experiment-based approach collects repeated measurements under more than one treatment level for the same individual from which some relevant information about the partial correlation between potential outcomes is available. The experiment designed for repeat measurements is known as N-1 trails(Shamseer et al., 2015; Araujo et al., 2016). Researchers could also design an auxiliary treatment ($W_i = 2$) and assign the extensive treatment to all individuals in the sample. Then the individual treatment effect $Y_i(1) - Y_i(0)$ can be evaluated by $[Y_i(2) - Y_i(0) - (Y_i(2) - Y_i(1))]$.

Both model-based and experiment-based approaches search for extra information to determine the partial correlation between potential outcomes. If such additional information is not available, the alternative is sensitivity analysis. After imputing the missing potential outcomes, one could evaluate individual treatment effects with various valid partial correlations and study the effect of partial correlations on the conclusions (Gadbury, Iyer & Allison, 2001).

3 Multiple imputation of multivariate incomplete variables

Datasets used for evaluating individual treatment effects by multiple imputation often have incomplete covariates and potential outcomes. We impute potential outcomes with joint modeling (JM) and covariates with fully conditional specification (FCS). Usually, one framework is used to generate all imputations, but Van Buuren (2018) highlighted that a blocked approach could be adopted to accommodate for hybrid versions of JM within FCS. We will now briefly introduce joint modelling, fully conditional specification and so-called hybrid imputation.

3.1 Joint modeling imputation (JM)

Joint modeling imputation assumes a model $p(Y^{mis}, Y^{obs} | \theta)$ for the complete data and a prior distribution $p(\theta)$ for the parameter θ. Joint modelling partitions the observed data into groups based on the missing pattern and imputes the missing data within each missing pattern according to corresponding predictive distribution. Under the assumption of ignorability, the parameters of the predictive distribution for different missing patterns are generated from the posterior joint distribution. Schafer (1997) proposed joint modelling methods for multivariate normal data, categorical data and mixed normal-categorical data. The joint modelling approach has solid theoretical properties (i.e., compatibility between the imputation and substantive models) while it lacks the flexibility of model specification.

3.2 Fully Conditional Specification (FCS)

In Fully Conditional Specification, we specify the distribution for each partially observed variable conditional on all other variables $P(Y(j)|Y(-j), X, \theta_j)$ and impute each missing variable iteratively. The FCS starts with naive imputations such as a random draw from the observed values. The tth iteration for the incomplete variable $Y(j)$ consists of the following draws:

$$\theta_j^t \sim f(\theta_j)f(Y^{obs}(j)|Y^{t-1}(-j), X, \theta_j)$$
$$Y^{mis(t)}(j) \sim f(Y^{mis}(j)|Y^t(-j), X, \theta_j^t),$$

where $f(\theta_j)$ is generally specified as a noninformative prior. After a sufficient number of iteration, typically with 5 to 10 iterations (van Buuren, 2018), the stationary distribution is achieved. The final iteration generates a single imputed dataset and the multiple imputations are created by applying FCS in parallel m times. Since FCS provides tremendous flexibility in specifying imputation models for multivariate partially observed data, FCS is now a widely accepted and popular MI approach (Van Buuren, 2007). Even while, FCS lacks a satisfactory theory and has a potential risk in incompatibility.

3.3 Block imputation

Block imputation combines the flexibility of FCS with the attractive theoretical properties of JM. A block consists of one or more variables. If the block has multiple variables, then we use multivariate imputation methods to impute those variables jointly. A simple example would be multiple imputation

of missing variables with quadratic effects $Y = \alpha + \beta_1 X + \beta_2 X^2 + \varepsilon$. In such a case, grouping the missing variable X and its corresponding square term X^2 within one block is of benefit to preserve the quadratic relationship (Vink, 2019). The joint modelling approach is the special case where all variables form one block, while the FCS approach treats each variable as a separate block.

When the imputation model of one variable is potentially incompatible, or its theoretical properties are not fully studied (i.e., whether the imputations based on the FCS correspond to drawing from a joint distribution), block imputation would merge that variable with other variables and apply the joint modeling imputation approach to that block. On the other hand, when the joint distribution of several missing variables is ambiguous, block imputation could use the FCS approach to impute each variable. In general, the apparent advantage of block imputation is the flexibility of model specification. However, block methods are hardly known or studied. While available in the `mice` software, the properties of block imputation have yet to be studied.

4 Specified partial correlation imputation

In this section we detail how blocked imputation can be used to impute missing outcomes with a given partial correlation between potential outcomes. We term the algorithm the imputation algorithm with specified partial correlation (SPC). Since the missing pattern of potential outcomes is somehow restrictive (see Figure 1) that is, no cases with completely observed potential

potential outcomes ⬛ covariates

Figure 1: Missingness mechanism of potential outcomes. The write represents observed value and the grey represents missing value. Without loss of generality, we assume covariates are completely observed.

outcomes, the imputation procedure follows three steps:

1. Estimating the marginal distribution of potential outcomes conditioning on pre-treatment variables.

2. Derive the multivariate density of potential outcomes by combining the marginal distribution of potential outcomes and the specified correlation between potential outcomes.

3. Impute the missing outcomes with the corresponding submodel obtained from the multivariate distribution.

Based on Rubin causal model (Imbens & Rubin 2015, Ch 8), it is plausible to assume a multivariate normal distribution for continuous potential outcomes. However, it is usually not valid to assume a joint distribution for the incomplete dataset. Applying fully conditional specification to impute covariates allows flexibility of imputation model specification. It is noticeable that SPC could also predict the posterior distribution of the individual treatment effect for units not in the experiment.

Our approach shares some similarities with statistical matching discussed by Moriarity & Scheuren (2003). For example, suppose there are two sample files, A and B. File A collects variables X and Y and file B collects variables X and Z. The purpose of statistical matching is to combine two files, A and B, into one file containing variables X, Y and Z. Rubin (1986) proposed a procedure of statistical matching with three steps: regression step, matching step and concatenation step. In the regression step, Rubin specified the correlation between variable X and Y to derive the joint distribution of (X, Y, Z) in two sample files. We develop this idea to evaluate the individual treatment effects and extend to multiple treatments condition.

Without loss of generality, let us assume that potential outcomes follow a multivariate normal distribution. We specify Bayesian linear models for two potential outcomes based on observed covariates.

$$Y(0) = \beta_0 X + \varepsilon_0, \varepsilon_0 \sim \mathcal{N}(0, \sigma_0^2) \tag{8}$$

$$Y(1) = \beta_1 X + \varepsilon_1, \varepsilon_1 \sim \mathcal{N}(0, \sigma_1^2). \tag{9}$$

Bayesian sampling draws β_0^*, β_1^*, σ_0^{*2}, σ_1^{*2} from their respective posterior

distribution. The Jeffrey's prior used and hence, the posterior distributions of σ_0^{*2} and σ_1^{*2} would be inverse χ^2 distribution:

$$\sigma_0^{*2} \sim \sum_{i=1}^{N_0} (Y_i(0) - \hat{\beta}_0 X_i)^2 \chi_{N_0-k}^{-2} \tag{10}$$

$$\sigma_1^{*2} \sim \sum_{i=1}^{N_1} (Y_i(1) - \hat{\beta}_1 X_i)^2 \chi_{N_1-k}^{-2}, \tag{11}$$

where $\hat{\beta}_0 = (X_0'X_0)^{-1}X_0'Y(0)$, $\hat{\beta}_1 = (X_1'X_1)^{-1}X_1'Y(1)$ and k is the number of covariates. The conditional distributions of β_0^* and β_1^* are multivariate normal:

$$\beta_0^* | \sigma_0^{*2} \sim \mathcal{N}(\hat{\beta}_0, \sigma_0^{*2}((X_0'X_0)^{-1})) \tag{12}$$

$$\beta_1^* | \sigma_1^{*2} \sim \mathcal{N}(\hat{\beta}_1, \sigma_1^{*2}((X_1'X_1)^{-1})). \tag{13}$$

Since there is no information relevant to the partial correlation between potential outcomes in the observed data, the posterior distribution of the partial correlation $\rho_{Y(0)Y(1)|X}$ equals the prior distribution specified by the user, who can select several numbers in the interval $[-1, 1]$ to investigate the sensitivity to $\rho_{Y(0)Y(1)|X}$. Finally, combined the marginal distribution of $Y(0)$ and $Y(1)$ with the specification of $\rho_{Y(0)Y(1)|X}$, the joint distribution of $Y^{mis}(0)$ and $Y^{mis}(1)$ is:

$$\begin{bmatrix} Y^{mis}(0) \\ Y^{mis}(1) \end{bmatrix} \sim \mathcal{N}\left[\begin{pmatrix} \beta_0^* x \\ \beta_1^* x \end{pmatrix}, \begin{pmatrix} \sigma_0^{*2} & \rho_{Y(0)Y(1)|X}\sigma_0^*\sigma_1^* \\ \rho_{Y(0)Y(1)|X}\sigma_0^*\sigma_1^* & \sigma_1^{*2} \end{pmatrix} \right] \tag{14}$$

and the distributions of $Y^{mis}(0)$ and $Y^{mis}(1)$ are:

$$Y^{mis}(0) \sim \mathcal{N}(\beta_0^* x + (Y(1) - \beta_1^* x)\rho_{Y(0)Y(1)|X}\sigma_0/\sigma_1, (1 - \rho_{Y(0)Y(1)|X}^2)\sigma_0^{*2})$$

(15)

$$Y^{mis}(1) \sim \mathcal{N}(\beta_1^* x + (Y(0) - \beta_0^* x)\rho_{Y(0)Y(1)|X}\sigma_1/\sigma_0, (1 - \rho_{Y(0)Y(1)|X}^2)\sigma_1^{*2}).$$

(16)

Comparing equations (15) and (16) to equations (8) and (9), it is evident that inclusion of the observed outcome may change the location of missing outcomes shifts slightly and the uncertainty is reduced when imputing missing outcomes under the specified correlation between potential outcomes. For the prediction of units out of trials, the reasonable values for outcomes under two treatments could be drawn from the joint distribution (14).

When generalizing to the multiple treatments condition $W = 0, 1, \ldots, w$, the marginal posterior distribution for potential outcomes would be:

$$Y(0)^* = \beta_0^* X + \varepsilon_0^*, \varepsilon_0^* \sim \mathcal{N}(0, \sigma_0^{*2})$$

$$Y(1)^* = \beta_1^* X + \varepsilon_1^*, \varepsilon_1^* \sim \mathcal{N}(0, \sigma_1^{*2})$$

$$\ldots$$

$$Y(w)^* = \beta_w^* X + \varepsilon_w^*, \varepsilon_w^* \sim \mathcal{N}(0, \sigma_w^{*2}),$$

where the values of $\beta_0^*, \beta_1^*, \ldots, \beta_w^*, \sigma_0^{*2}, \sigma_1^{*2}, \ldots, \sigma_w^{*2}$ draw from their respective Bayesian posterior distribution. If $\sigma_0^{*2}, \ldots, \sigma_w^{*2}$ are unrestricted, with pairwise specification of partial correlation between potential outcomes, the

joint distribution of $Y^{mis}(0)$, $Y^{mis}(1)$, ..., $Y^{mis}(w)$ is:

$$\begin{bmatrix} Y^{mis}(0) \\ Y^{mis}(1) \\ \ldots \\ Y^{mis}(w). \end{bmatrix} \sim \mathcal{N}(\mathcal{M}, \Sigma), \tag{17}$$

where $\mathcal{M} = (\beta_0^* x, \beta_1^* x, \ldots, \beta_w^* x)^{\mathsf{T}}$. The covariance matrix Σ must be positive semi-definite:

$$\Sigma = \begin{pmatrix} \sigma_0^{*2} & \rho_{Y(0)Y(1)|X}\sigma_0^*\sigma_1^* & \cdots & \rho_{Y(0)Y(w)|X}\sigma_0^*\sigma_w^* \\ \rho_{Y(1)Y(0)|X}\sigma_1^*\sigma_0^* & \sigma_1^{*2} & \cdots & \rho_{Y(1)Y(w)|X}\sigma_1^*\sigma_w^* \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{Y(w)Y(0)|X}\sigma_w^*\sigma_0^* & \rho_{Y(w)Y(1)|X}\sigma_w^*\sigma_1^* & \cdots & \sigma_w^{*2} \end{pmatrix}. \tag{18}$$

Draws of missing outcomes for units under different treatments could be derived from the joint distribution based on the property of conditional distribution for the multivariate normal distribution. For instance, with units under control treatment $W = 0$, the distribution of missing outcomes $Y^{mis}(-0) = (Y^{mis}(1), \ldots, Y^{mis}(w))$ would be $Y^{mis}(-0) \sim \mathcal{N}((\beta_1^* x, \ldots, \beta_w^* x)^{\mathsf{T}} + \Sigma_{0-0}\Sigma_{-0-0}^{-1}(Y_0 - \beta_0^* x), \Sigma_{00} - \Sigma_{0-0}\Sigma_{-0-0}^{-1}\Sigma_{-00})$, where

$$\begin{bmatrix} \Sigma_{00} & \Sigma_{0-0} \\ \Sigma_{-00} & \Sigma_{-0-0} \end{bmatrix}, \tag{19}$$

is the partition of Σ: $\Sigma_{00} = \sigma_0^{*2}$, $\Sigma_{0-0} = \Sigma_{-00}^{\mathsf{T}} = (\rho_{Y(0)Y(1)|X}\sigma_0^*\sigma_1^*, \ldots, \rho_{Y(0)Y(w)|X}\sigma_0^*\sigma_w^*)$

and

$$\Sigma_{-0-0} = \begin{pmatrix} \sigma_1^{*2} & \cdots & \rho_{Y(1)Y(w)|X}\sigma_1^*\sigma_w^* \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{Y(w)Y(1)|X}\sigma_w^*\sigma_1^* & \cdots & \sigma_w^{*2} \end{pmatrix}. \qquad (20)$$

One could use the sweep operator for rapid calculation of the parameters for imputation models of missing outcomes (Goodnight, 1979).

5 Simulation study

We evaluate the performance of SPC at both the individual level (i.e. the individual treatment effect) and the aggregate level (i.e. the average treatment effect). For individual causal inference, we study the mean and mean absolute differences between the 'true' and imputed individual treatment effects, together with posterior distributions of individual treatment effects. For average causal inference, we analyse biases and confidence interval coverages of the estimated parameters in the distribution of the potential outcomes. We perform a sensitivity analysis to the multiple imputation approach with three different values for the partial correlation between potential outcomes: $\rho = 0, 0.73$ or 0.99, which correspond to, respectively, a conditional independent correlation assumption, the correct partial correlation and a constant treatment effect condition.

We compare the performance of SPC to the targeted learning approach by Van der Laan & Rose (2011). Targeted learning is an alternative for estimating individual treatment effects. The idea is to estimate the data-

generating distribution P_0 and then update the initial estimation to make an optimal bias-variance tradeoff for the scientific interest $\Psi(P_0)$. To estimate individual treatment effects, we define the average treatment effect as the scientific interest:

$$\Psi(P_0) = E[E(Y_i(1) \mid X) - E(Y_i(0) \mid X)]. \tag{21}$$

The targeted learning consists of three steps of analysis: 1) definition of the data-generating model and the scientific interest $\Psi(P_0)$, 2) super learning for initial prediction of $\Psi(P_0)$ and 3) targeted maximum likelihood estimation for $\Psi(P_0)$ (Van der Laan & Rose, 2011). Specifically, before estimation, it is necessary to define a set of possible probability distributions of observed data and identify a collection of causal assumptions (i.e., the ignorable assignment mechanism and the stable unit treatment value assumption) to the identification of the correct model. With the definition of the model, one could apply the super learner to derive an initial estimation for the distribution of potential outcomes $\hat{P}_0$. The super learner first selects a library of candidate algorithms and a risk function and then applies the validation set approach to calculate the average risk for each algorithm. The optimal algorithm with the smallest average risk is used to produce the initial predicted distribution of potential outcomes. The candidate algorithms could be parametric(i.e., general linear model), non-parametric (i.e., random forest), or even a weighted combination of statistical algorithms.

After the initial estimation of predicted potential outcomes, one could define the targeted maximum likelihood estimation (TMLE) for scientific in-

terest $\Psi(P_0)$. The TMLE step reduces the bias in the estimation of $\Psi(P_0)$ if the initial estimation $\Psi(\hat{P}_0)$ is inconsistent. This is accomplished by exploiting information in the treatment assignment mechanisms to adjust the initial estimations. Generally, the adjustment is an iterative procedure. However, when the scientific interest is the average treatment effect, convergence is achieved in one step. More details are provided in Gruber & van der Laan (2011).

While targeted learning is a machine learning approach aimed at estimating the average treatment effect, it involves calculating the missing potential outcomes and can therefore also be used to identify individual treatment effects. However, unlike SPC, the targeted learning fits the distribution of the missing outcome only based on the covariates, which assumes conditional independence between potential outcomes. In the simulation study, we aim to show the relevance of specifying the correlation between potential outcomes when specifying the analysis model of potential outcomes.

5.1   Simulation conditions

We design the two potential outcomes $Y(0)$ and $Y(1)$ as well as one baseline covariate X. The data is generated with a multivariate normal distribution:

$$
\begin{pmatrix} Y_0 \\ Y_1 \\ X \end{pmatrix} \sim \mathcal{N} \left[\begin{pmatrix} 0 \\ 1 \\ 2 \end{pmatrix}, \begin{pmatrix} 1 & 0.8 & 0.5 \\ 0.8 & 1 & 0.5 \\ 0.5 & 0.5 & 1 \end{pmatrix} \right]
$$

Because the marginal correlation between potential outcomes is 0.8, the corresponding partial correlation is 0.73.

$$
\begin{aligned}
\rho_{y_0 y_1 | x} &= \frac{\rho_{y_0 y_1} - \rho_{y_0 x} \rho_{y_1 x}}{\sqrt{1 - \rho_{y_0 x}^2} \sqrt{1 - \rho_{y_1 x}^2}} \\
&= \frac{0.8 - 0.25}{0.75} \\
&\approx 0.73,
\end{aligned}
\tag{22}
$$

A total of $N = 5000$ independent and identically distributed cases are generated. The first 2500 cases have only information for Y_0 and the remaining cases have only information on Y_1. One thousand repetitions of the simulation are produced for average causal inference. While, for individual causal inference, we derive the posterior distributions of imputed outcomes from twenty imputed datasets. For reasons of brevity, we only include one pre-treatment covariate. However, it is straightforward to extend the methodology to situations with more covariates and even mixtures of continuous and categorical predictors.

5.2   Results

5.2.1   Individual causal inference

We define bias as the mean difference between *true* and estimated values over twenty imputed datasets. Figure 2 shows the distribution of the bias for all four strategies, and the corresponding location and scale are displayed in Table 1.

Overall, SPC with three different partial correlations and the targeted learning all yield unbiased estimates of the average treatment effect. How-
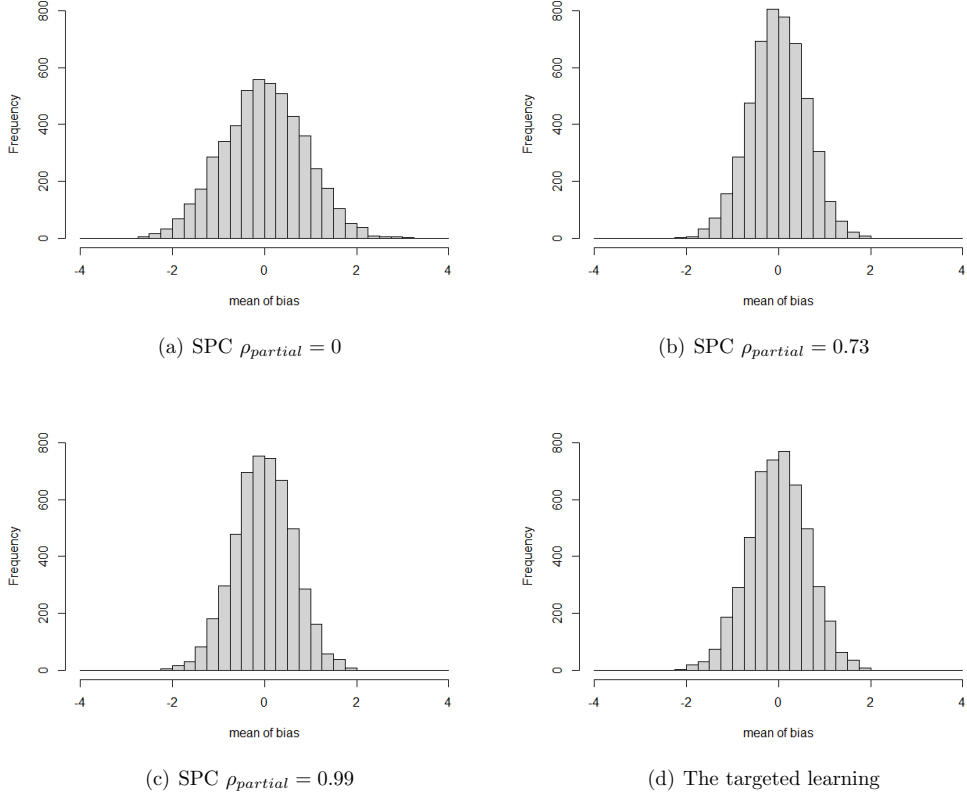
21

(a) SPC $\rho_{partial} = 0$

(b) SPC $\rho_{partial} = 0.73$

(c) SPC $\rho_{partial} = 0.99$

(d) The targeted learning

Figure 2: Histrogram plots of the bias of estimate the individual treatment effects for all four strategies.

$\rho_{Y(0)Y(1)\mid X}$	Mean bias	Variance of bias
SPC $\rho_{partial} = 0$	-0.012	0.791
SPC $\rho_{partial} = 0.73$	-0.007	0.363
SPC $\rho_{partial} = 0.99$	-0.013	0.398
The targeted learning	-0.006	0.401

Table 1: The location and the scale for the bias of estimate the individual treatment effects for all four strategies.

ever, in terms of the scale, The SPC with the *correct* partial correlation has a minor variance. The closer the specified partial correlation is to the partial correlation in the *true* data generating model, the smaller bias and variance
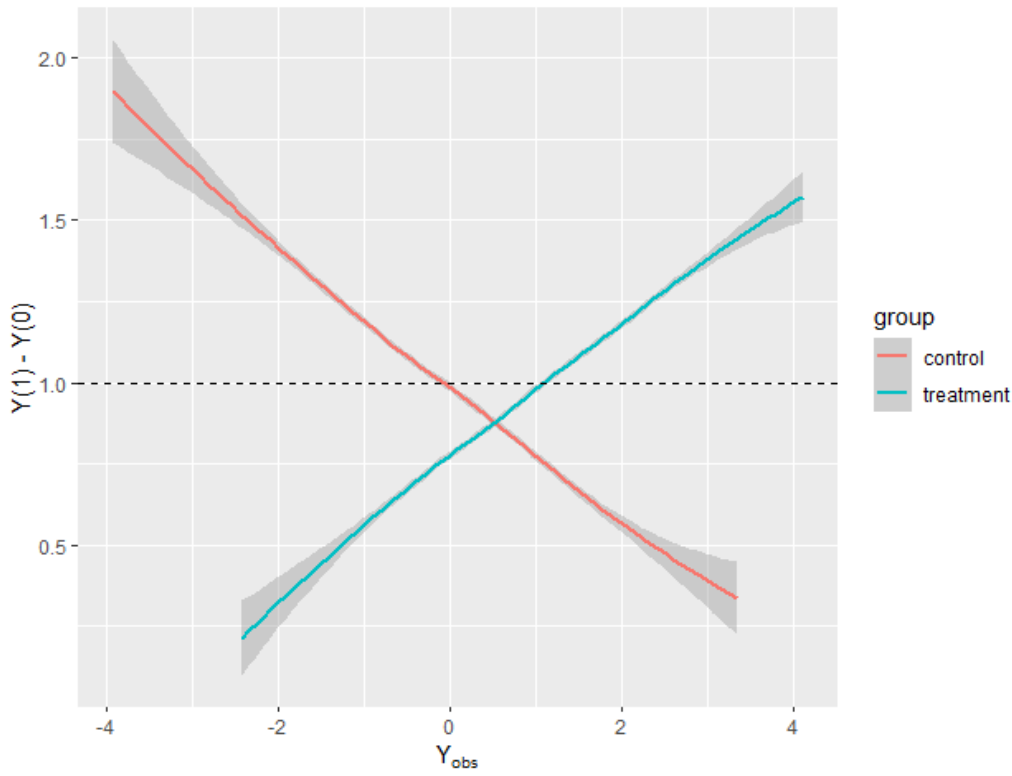
Figure 3: The plot shows mean individual treatment effects $(Y(1) - Y(0))$ in both treatment and control group when partial correlation is defined as 0.73. The oucome $Y(0)$ is observed in the control group and the oucome $Y(1)$ is observed in the treatment group. The dashed line represents the average treatment effect.

can be expected. Although it is difficult to produce accurate estimates of the individual treatment effect for units at the tail in Fig 3, we still have a large proportion of the estimated individual treatment effect with negligible biases. Since the variance of the bias equals the partial variance of the potential outcomes, we could include more explanatory variables to increase the accuracy of the imputation of missing outcomes and hence the prediction of individual treatment effect. On the other hand, when the specified partial correlation deviates from the *true* value, more variance would appear

because of the differences between the *true* distribution and the estimated distribution for each missing outcome.
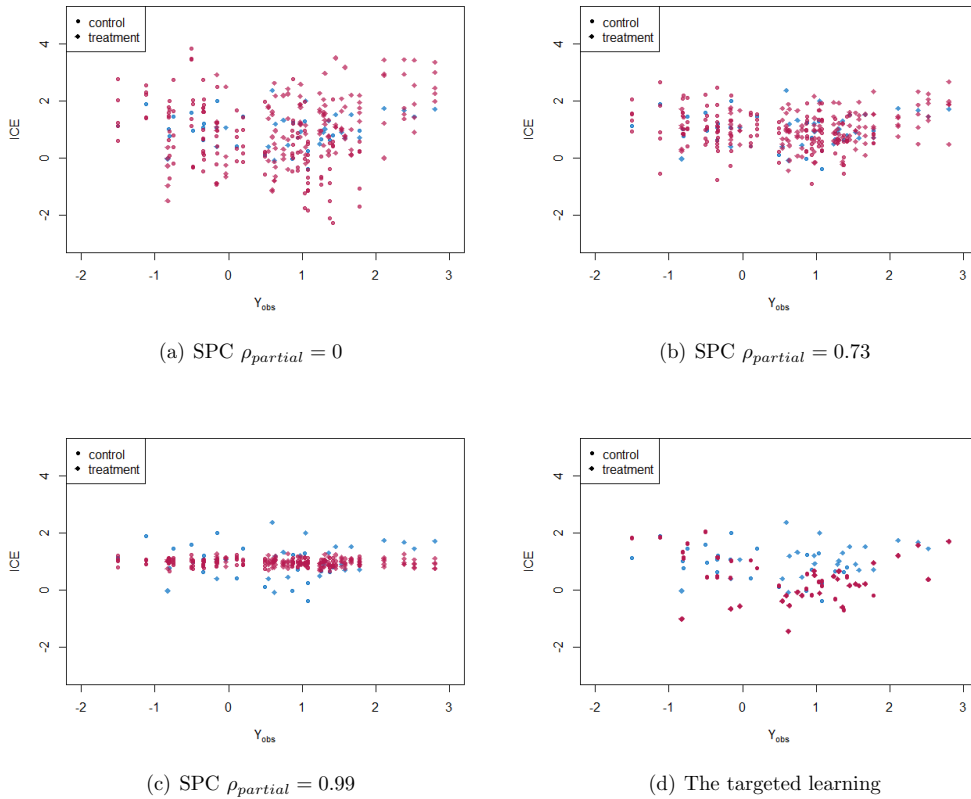

(a) SPC $\rho_{partial} = 0$

(b) SPC $\rho_{partial} = 0.73$

(c) SPC $\rho_{partial} = 0.99$

(d) The targeted learning

Figure 4: Stripplot of m = 5 of observed (blue) and imputed (red) data for individual treatment effects with selected cases.

Figure 4 shows posterior distributions of individual treatment effects for selected cases($i = 100, 200, \ldots, 5000$). When the partial correlation is specified correctly, the imputations look plausible: imputed ITE covers the *true* ITE for almost every case, and the variance of ITE for each individual is smaller than the case under the independent conditional correlation assumption. With homogeneous treatment effect assumption, i.e.,

$\rho_{Y(0)Y(1) \mid X} = 0.99$, the imputed individual treatment effects are biased towards the average treatment effect. For targeted learning, uncertainty about the missing outcomes is not estimated.

Furthermore, we evaluate the imputations with all possible positive partial correlation (range from 0 to 1) by mean distance between the *true* and the mean estimated individual treatment effects and the rate of the posterior distribution of imputed outcome cover the *true* value. Fig 5 shows that the violation of the homogeneous treatment effect assumption leads to extremely poor coverage. If we specified the partial correlation closed to the *true* value, the imputed outcomes would be more accurate (see Fig 6). Fig 6 also highlight the mean distance calculated by the targeted learning, which implies the implicit assumption of conditional independence between potential outcomes.

The SPC approach derives the distribution of individual treatment effects, which provides more information on treatment recommendations. For instance, with a small individual treatment effect, it is possible to estimate the probability of a positive treatment effect from the distribution of individual treatment effects.

5.2.2 Average causal inference

In this section, we investigate whether we could provide valid inferences for the distribution of the potential outcomes. In addition, we are interested in the biases and the coverage of nominal 95% confidence intervals of all parameters related to the potential outcomes. Table 2 shows all statistics relevant to the potential outcomes. Statistics involving only one potential
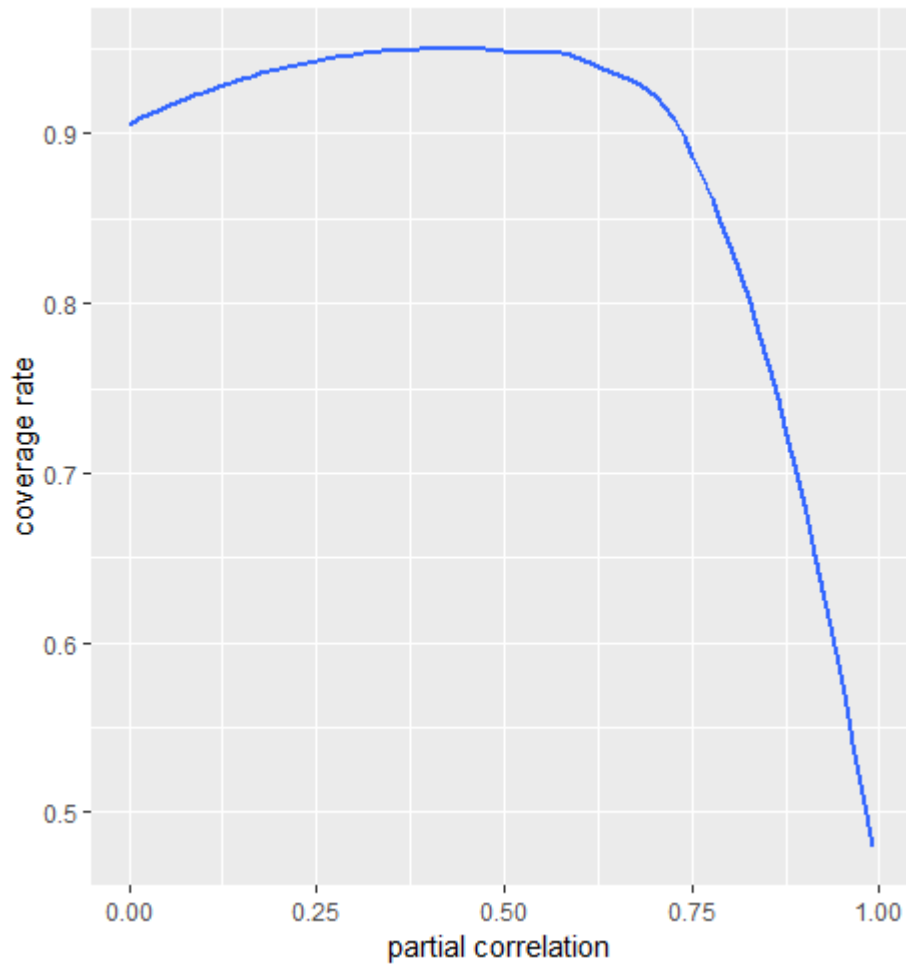
Figure 5: Coverage rate of the posterior distribution of estimated individual treatment effects

outcome are unbiased and have valid coverage rates, which means that even with incorrect specified partial correlation, we could derive plausible marginal distribution of potential outcomes. Since the partial correlation is set before imputation and there is no information about the correlation between potential outcomes in the data, we get valid inference for the marginal correlation between potential outcomes only when we specified the partial correlation correctly.
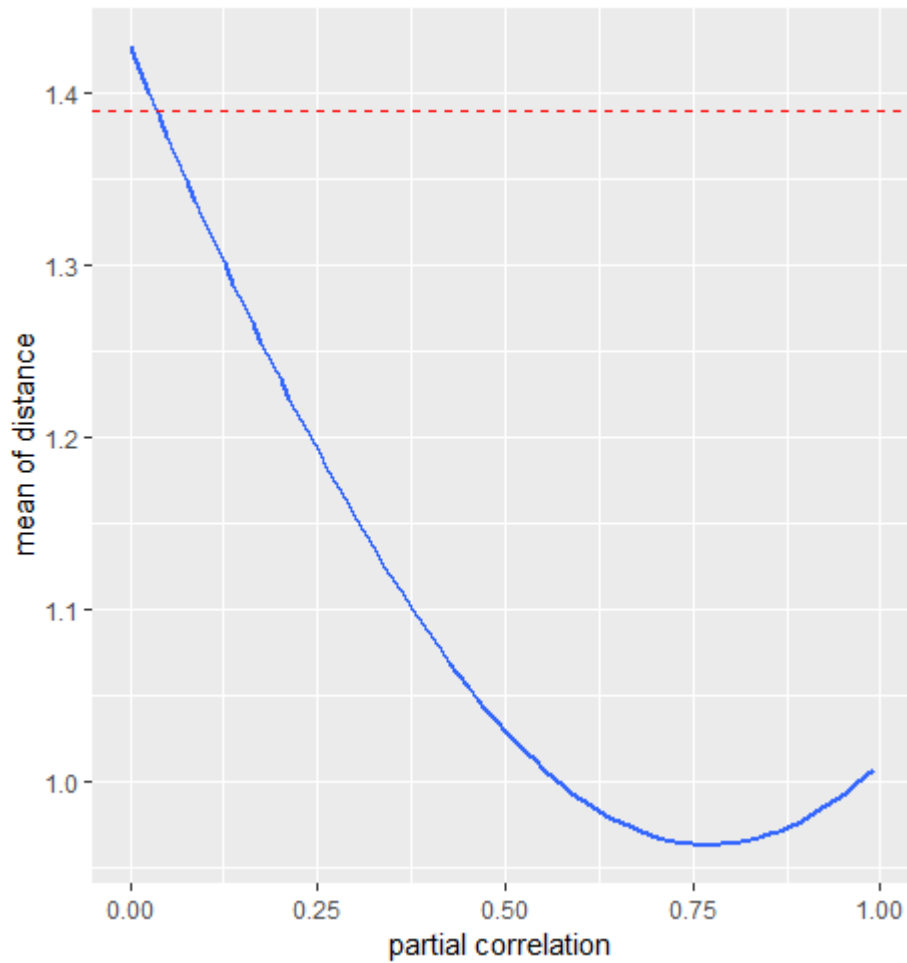
26

Figure 6: Mean distance between the *true* and the mean estimated individual treatment effects. The red dashed line represents the mean distance calculated by the targeted learning approach.

6   Application

We apply the SPC algorithm to evaluate the effects of two different therapies on slowing the progression of HIV disease. The data comes from a study comparing the effects of four therapies (zidovudine alone, didanosine alone, zidovudine plus didanosine and zidovudine and zalcitabine) on preventing the deterioration of disease in adults with HIV-1 infected pa-

Method	Truth	Est	Cover
SPC $\rho_{partial} = 0$			
$E(y_0)$	0.0	0.00	0.95
$E(y_1)$	1.0	1.00	0.94
$Var(y_0)$	1.0	1.00	0.94
$Var(y_1)$	1.0	1.00	0.94
$Cov(y_0, y_1)$	0.8	0.25	0.00
$Cov(y_0, x)$	0.5	0.50	0.94
$Cov(y_1, x)$	0.5	0.50	0.95
SPC $\rho_{partial} = 0.73$			
$E(y_0)$	0.0	0.00	0.97
$E(y_1)$	1.0	1.00	0.95
$Var(y_0)$	1.0	1.00	0.94
$Var(y_1)$	1.0	1.00	0.95
$Cov(y_0, y_1)$	0.8	0.80	1.00
$Cov(y_0, x)$	0.5	0.50	0.95
$Cov(y_1, x)$	0.5	0.50	0.94
SPC $\rho_{partial} = 0.99$			
$E(y_0)$	0.0	0.00	0.95
$E(y_1)$	1.0	1.00	0.95
$Var(y_0)$	1.0	1.00	0.95
$Var(y_1)$	1.0	1.00	0.94
$Cov(y_0, y_1)$	0.8	0.99	0.00
$Cov(y_0, x)$	0.5	0.50	0.95
$Cov(y_1, x)$	0.5	0.50	0.95

Table 2: Parameter estimates for SPC with different partial correlations

tients (Hammer et al., 1996). For simplicity, we name them treatment A(zidovudine alone), B(didanosine alone), C(zidovudine plus didanosine) and D(zidovudine and zalcitabine). The data named ACTG175 is accessible in package `speff2trial` in R. We restrict our analyses to treatments A and B and perform out-of-sample prediction of hypothetical effect of treatments A and B for the remaining patients that were allocated to treatments C and D. Hammer et al. concluded that treatment with didanosine is superior to treatment with zidovudine. However, the overall treatment effect was found

to be insufficient to recommend the therapy to a patient, a situation that is common in many medical interventions.

A total of 693 HIV-1 infected adults with CD4 cell counts in the range of 200 to 500 per cubic millimetre were randomised into the control (N = 316) and treatment (N = 377) group, while 670 patients were treated as out-of-sample. Fifteen baseline covariates are included which assess gender, age, weight, Karnofsky score, risk factors, prior antiretroviral therapy, CD4 cell count, and CD8 T cell count. We are interested in the number of days until the first occurrence of: 1) a decline in CD4 cell count of at least 50 2) an event indicating progression to AIDS, or 3) death. The larger number of days yields a more beneficial treatment effect. The individual treatment effect is defined as the number of days under treatment B minus the number of days under treatment A. We select the value of partial correlation as 0 and 0.7 to perform the sensitivity analysis so that the result yields distinct differences.
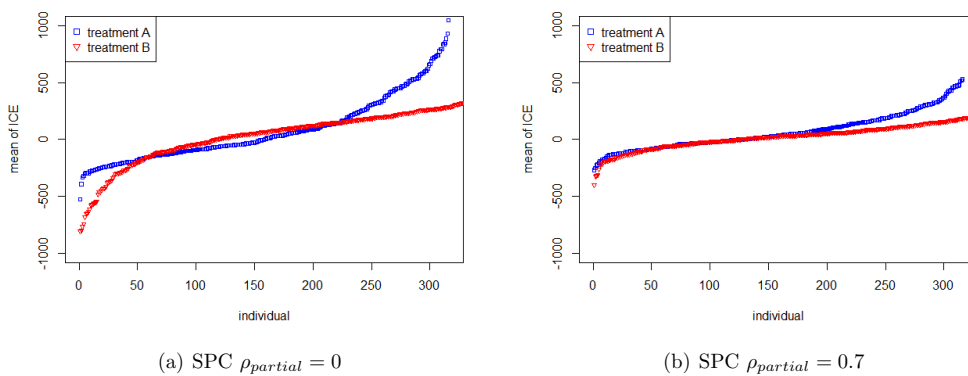

(a) SPC $\rho_{partial} = 0$ (b) SPC $\rho_{partial} = 0.7$

Figure 7: Means of individual treatment effects for patients under treatment A and treatment B, which are arranged into ascending order.

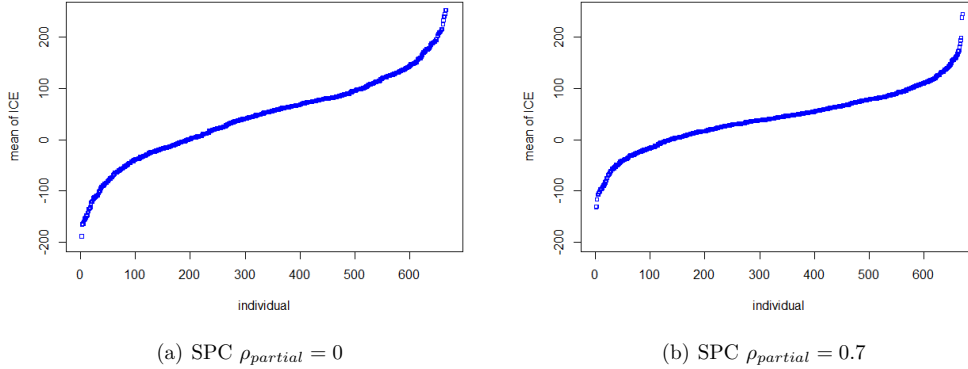(a) SPC $\rho_{partial} = 0$ (b) SPC $\rho_{partial} = 0.7$

Figure 8: Means of individual treatment effects for out-of-sample patients under treatment C and D, which are arranged into ascending order.

Figure 7 shows the results of individual treatment effects under treatment A and treatment B group with partial correlation specification 0 and 0.7. As expected, this approach could detect the heterogeneity of treatment effects. A large proportion of patients in the sample, whose individual treatment effects are larger than 0, are recommended to receive a treatment regimen with didanosine. However, treatment with zidovudine still yields greater clinical benefit for a fraction of units, whose individual treatment effects are smaller than 0. Since all covariates are balanced under two groups, the distributions of expected individual treatment effects under two treatments (A and B) are more similar when specifying a 0.7 partial correlation. Furthermore, the range of expected value of individual treatment effects is smaller, with a partial correlation of 0.7. The larger the partial correlation we set, the more convinced that all effect modifiers are included, and effects are identical across persons. Figure 8 shows variability in individual effects in out-of-sample patients, which implies that our method could also be applied to a
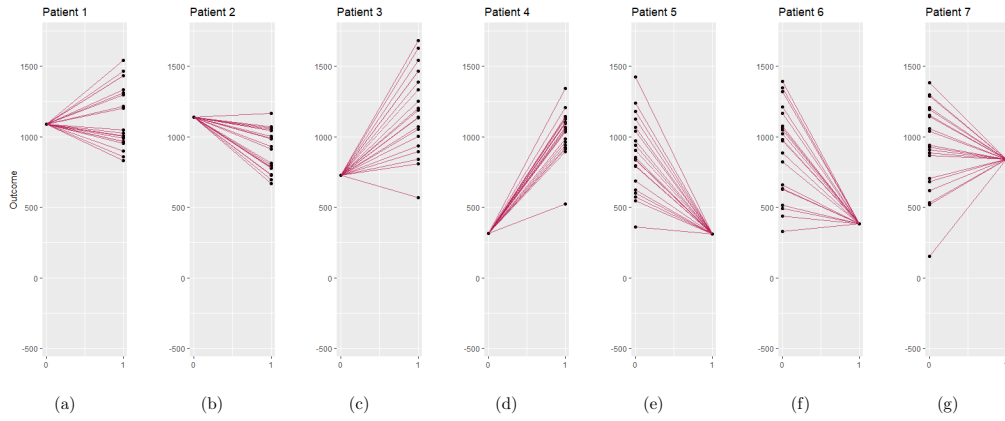
prediction scenario.

Figure 9: Fan plot of Observed and imputed (m = 20) outcomes under treatment A (0) and B (0). The partial correlation is 0.
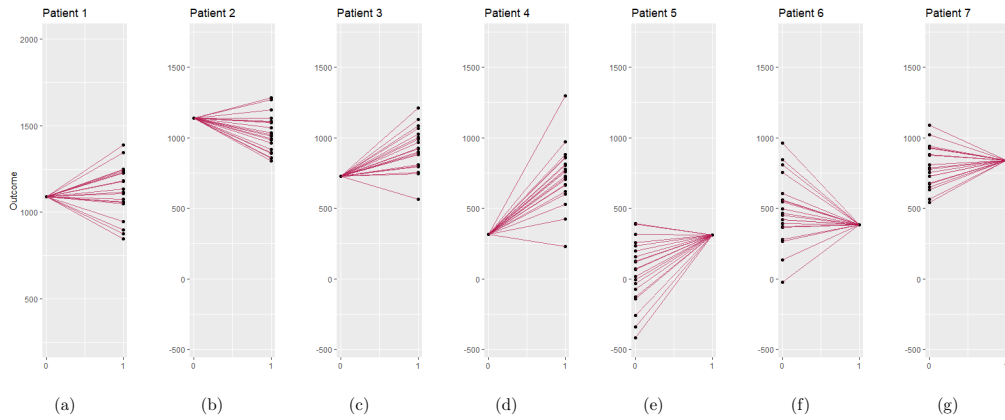

Figure 10: Fan plot of Observed and imputed (m = 20) outcomes under treatment A (0) and B (0). The partial correlation is 0.7.

Figure 9 and 10 display imputations by selected patients for two different values of partial correlation, 0 and 0.7. Each panel contains the observed outcome for the patient and $m = 20$ imputed values for the missing outcome. The imputed outcomes are sensitive to different values of partial correlation.

Patient 5 benefits from treatment A when the partial correlation equals 0, while we derive the opposite conclusion when specifying the partial correlation as 0.7. The location of imputed outcomes is different under various partial correlations, and the scale shrinks when the partial correlation tends to 1.

7 Discussion

We propose a multiple imputation approach to replace missing outcomes with plausible values to estimate the individual treatment effect. Treatment assignment is currently steered by the average treatment effect. This may lead to suboptimal individual treatment decisions because the treatment effects are assumed to be homogeneous. We have demonstrated that the proposed SPC algorithm allows for the imputation of heterogeneous treatment effects under a given partial correlation.

The sensitivity analysis of the partial correlation between potential outcomes in section 5.2 demonstrates that different values of the partial correlation yield similar results in terms of marginal distributions of potential outcomes and the average treatment effect. However, the closer the specified partial correlation is to the *true* value, the less biased the estimated individual treatment effect will be. Since one cannot obtain information about the partial correlation from the observed data, the determination of the partial correlation should be set to a plausible range based on previous investigations or expert knowledge. In addition, it may be useful to perform a sensitivity analysis to see how imputed outcomes differ with different partial correla-

tions.

An advantage of our multiple imputation approach to individual treatment effects is that it provides an estimate of the uncertainty of the imputed outcomes and hence, of the individual treatment effects. One could obtain the posterior distribution of the individual treatment effects for a unit from multiply imputed datasets, from which we can learn the probability of benefit from the treatment at the individual level. Since we incorporate the partial correlation when imputing, researchers could apply complete-data analyses to explore potential variables. This accounts for residual heterogeneity of treatment effects or additional effect modifiers.

In our illustration of the SPC algorithm, we applied Bayesian imputation under the normal linear model. It is possible to use other imputation techniques (parametric or non-parametric imputation methods). The behaviour of such methods has not yet been studied. Since SPC is a hybrid of FCS and JM, it will provide valid inferences on data that are missing at random. Another useful property of SPC is that under the assumption of ignorable treatment assignment, researchers can skip explicit modelling of the probability of assignment.

In the application study, we focus on the comparison of treatments A and B. It is possible to generalise the multiple treatment comparison (treatment A, B, C, and D). By imputing unobserved outcomes, we could then recommend the optimal treatment to each unit among four treatments. One could benefit from our method when performing an experiment that has been investigated on a different population. Some proven effect modifiers may be difficult to collect when performing the same experiment in other regions

33

or countries. In such a case, the inference would include the heterogeneity of treatment effects explained by the uncollected factors by specifying a reasonable partial correlation.

This is an initial study on MI to the individual treatment effect. The simulation study used a basic randomised trial with a correctly specified imputation model. Further work should be done to extend discrete and semi-continuous outcomes. Another challenge is to develop imputation techniques for studies that collect post-treatment variables. All in all, we believe that our methodology for incorporating the partial correlation represents an important advance in estimating individual treatment effects. We hope that our method may attribute to the growing body of work on personalised statistics and individual treatment effects.

References

Araujo, A., Julious, S., & Senn, S. (2016). Understanding variation in sets of n-of-1 trials. *PLoS One*, *11*(12), e0167167.

Ding, P., Feller, A., & Miratrix, L. (2019). Decomposing treatment effect variation. *Journal of the American Statistical Association*, *114*(525), 304–317.

Ding, P., Li, F., et al. (2018). Causal inference: A missing data perspective. *Statistical Science*, *33*(2), 214–237.

Djebbari, H., & Smith, J. (2008). Heterogeneous impacts in progresa. *Journal of Econometrics*, *145*(1-2), 64–80.

Gadbury, G. L., Iyer, H. K., & Allison, D. B. (2001). Evaluating subject-treatment interaction when comparing two treatments. *Journal of Biopharmaceutical Statistics*, *11*(4), 313–333.

Goodnight, J. H. (1979). A tutorial on the sweep operator. *The American Statistician*, *33*(3), 149–158.

Gruber, S., & Van der Laan, M. J. (2011). tmle: An r package for targeted maximum likelihood estimation.

Hammer, S. M., Katzenstein, D. A., Hughes, M. D., Gundacker, H., Schooley, R. T., Haubrich, R. H., . . . others (1996). A trial comparing nucleoside monotherapy with combination therapy in hiv-infected adults with cd4 cell counts from 200 to 500 per cubic millimeter. *New England Journal of Medicine*, *335*(15), 1081–1090.

Heckman, J. J. (2005). The scientific model of causality. *Sociological methodology*, *35*(1), 1–97.

Heckman, J. J., Humphries, J. E., Urzua, S., & Veramendi, G. (2010). The effects of schooling on labor market and health outcomes. *Unpublished Manuscript. University of Chicago, Department of Economics*.

Hernan, M. A., & Robins, J. M. (2010). *Causal inference.* CRC Boca Raton, FL.

Imbens, G. W., & Rubin, D. B. (2015). *Causal inference in statistics, social, and biomedical sciences.* Cambridge University Press.

Lamont, A., Lyons, M. D., Jaki, T., Stuart, E., Feaster, D. J., Tharmaratnam, K., ... Van Horn, M. L. (2018). Identification of predicted individual treatment effects in randomized clinical trials. *Statistical methods in medical research*, *27*(1), 142–157.

Moriarity, C., & Scheuren, F. (2003). A note on rubin's statistical matching using file concatenation with adjusted weights and multiple imputations. *Journal of Business & Economic Statistics*, *21*(1), 65–73.

Rässler, S. (2012). *Statistical matching: A frequentist theory, practical applications, and alternative bayesian approaches* (Vol. 168). Springer Science & Business Media.

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of educational Psychology*, *66*(5), 688.

Rubin, D. B. (1986). Statistical matching using file concatenation with adjusted weights and multiple imputations. *Journal of Business & Economic Statistics*, *4*(1), 87–94.

Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys.* New York: John Wiley and Sons.

Rubin, D. B. (2005). Causal inference using potential outcomes: Design, modeling, decisions. *Journal of the American Statistical Association*, *100*(469), 322–331.

Shamseer, L., Sampson, M., Bukutu, C., Schmid, C. H., Nikles, J., Tate, R., ... others (2015). Consort extension for reporting n-of-1 trials (cent) 2015: Explanation and elaboration. *Bmj*, *350*, h1793.

Smink, W. (2016). *Towards estimation of individual causal effects: The use of a prior for the correlation between potential outcomes* (Unpublished master's thesis).

Van Buuren, S. (2007). Multiple imputation of discrete and continuous data by fully conditional specification. *Statistical methods in medical research*, *16*(3), 219–242.

van Buuren, S. (2018). *Flexible imputation of missing data, second edition.* Chapman and Hall/CRC. doi: 10.1201/9780429492259

Van der Laan, M. J., & Rose, S. (2011). *Targeted learning: causal inference for observational and experimental data.* Springer Science & Business Media.

Vink, G., & van Buuren, S. (2019, December). *Hybrid imputation.* Retrieved 22.01.2020, from `https://www.gerkovink.com/London2019/`

Westreich, D., Edwards, J. K., Cole, S. R., Platt, R. W., Mumford, S. L., & Schisterman, E. F. (2015). Imputation approaches for potential outcomes in causal inference. *International journal of epidemiology*, *44*(5), 1731–1737.