



Minerva Access is the Institutional Repository of The University of Melbourne

Author/s:

Xu, J;De Silva, AP;Lee, KJ;Mahar, RK;Simpson, JA

Title:

Optimising dynamic treatment regimens using sequential multiple assignment randomised trials data with missing data

Date:

2025-12-01

Citation:

Xu, J., De Silva, A. P., Lee, K. J., Mahar, R. K. & Simpson, J. A. (2025). Optimising dynamic treatment regimens using sequential multiple assignment randomised trials data with missing data. *BMC Medical Research Methodology*, 25 (1), pp.162-. <https://doi.org/10.1186/s12874-025-02595-1>.

Persistent Link:

<https://hdl.handle.net/11343/361905>

License:

[CC BY](#)

RESEARCH

Open Access



Optimising dynamic treatment regimens using sequential multiple assignment randomised trials data with missing data

Jessica Xu¹, Anurika P. De Silva¹, Katherine J. Lee^{2,3}, Robert K. Mahar^{1,2†} and Julie A. Simpson^{1,4*†}

Abstract

Dynamic treatment regimens are commonly used for patients with chronic or progressive medical conditions. Sequential multiple assignment randomised trials (SMARTs) are studies used to optimise dynamic treatment regimens by repeatedly randomising participants to treatments. Q-learning, a stage-wise regression-based method used to analyse SMARTs, uses backward induction to compare treatments administered as a sequence. Missing data is a common problem in randomised trials and can be complex in SMARTs given the sequential randomisation. Common methods for handling missing data such as complete case analysis (CCA) and multiple imputation (MI) have been widely explored in single-stage randomised trials, however, the only study that explored these methods in SMARTs did not consider Q-learning. We evaluated the performance of CCA and MI on the estimation of Q-learning parameters in a SMART. We simulated 1000 datasets of 500 participants, based on a SMART with two stages, under different missing data scenarios defined by missing directed acyclic graphs (m-DAGS), percentages of missing data (20%, 40%), stage 2 treatment effects, and strengths of association with missingness in stage 2 treatment, patient history and outcome. We also compared CCA and MI using retrospective data from a longitudinal smoking cessation SMART. When there was no treatment effect at either stage 1 or 2, we observed close to zero absolute bias in the stage 1 treatment effect and similar empirical standard errors for CCA and MI under all missing data scenarios. When all participants had a relatively large stage 2 treatment effect, we observed minimal bias from both CCA and MI, with slightly greater bias for MI. Empirical standard errors were higher for MI compared to CCA under all scenarios except for when data were missing not dependent on any variables. When the stage 2 treatment effect varied between participants and data were missing dependent on other variables (for example, stage 1 responder status missing dependent on stage 1 treatment and baseline variables), we observed greater bias for MI when estimating the stage 1 treatment effect, which increased with the percentage missingness, while the bias for CCA remained minimal. Resulting empirical standard errors were lower or similar for MI compared to CCA under all missing data scenarios. Results showed that for a two-stage SMART, MI failed to capture the differences between treatment effects when the stage 2 treatment effect varied between participants.

Keywords Sequential multiple assignment randomised trials, Missing data, Q-learning, Multiple imputation

[†]Robert K. Mahar and Julie A. Simpson Joint senior authors.

*Correspondence:

Julie A. Simpson
julieas@unimelb.edu.au

Full list of author information is available at the end of the article



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Background

Clinicians must often make multi-stage dynamic decisions when treating patients with chronic or progressive diseases such as cancer [1] or mental health disorders [2]. Dynamic treatment regimens (DTRs) are a set of sequential treatment rules commonly used to guide these treatment decisions [3]. Sequential multiple assignment randomised trials (SMARTs) are clinical studies with multiple stages of randomisation [4] that can be used to estimate the value of DTRs and thus optimise sequential treatment decisions.

Missing data are inevitable in many randomised controlled trials (RCTs) and primarily occur in the outcome variable(s). In SMARTs, missing data can be more complicated due to the multiple stages of randomisation where data collection in subsequent stages depends on treatment allocation and participant outcomes in previous stages. For example, in a two-stage SMART, if a participant drops out during stage 1, then no data would be collected at stage 2 including treatment allocation and the outcome.

One common method for analysing SMART data is Q-learning. Q-learning is a stage-wise method that uses regression models to estimate the quality of a treatment decision at each treatment stage while appropriately accounting for previous decisions, participant outcome(s), and other baseline variables [5]. An advantage of using Q-learning to analyse SMART data is that it accounts for delayed effects from previous stages by starting the analysis from the final stage and working backwards.

Selecting an appropriate statistical method for handling the missing data that is based on pre-specified missing data assumptions is important [6]. Common methods to handle missing data include complete case analysis (CCA) and multiple imputation (MI). Only participants who have complete data for all analysis variables are included in a CCA [7]. MI is a two-stage process that repeatedly imputes missing values based on the observed data, then performs the target analysis on each imputed dataset, and combines the resulting parameter estimates and standard errors while accounting for the uncertainties associated with the missing values [8]. Although the performance of CCA and MI have been widely explored in conventional, single stage randomised trials [9], only a single study, led by Shortreed et al. [10], has explored how well these methods perform in a SMART setting when weighted regression (weights used to adjust for unequal randomisation probabilities) was used to estimate the mean response for each treatment regime. Shortreed et al. identified multiple challenges associated with using imputation methods for handling missing data in SMARTs [10], and pointed to the need for further

research to understand how MI methods perform when the underlying missing data assumptions are violated.

Because Q-learning is a complex analytic approach using backward stepwise regression of the multiple stages of SMART designs, it complicates how missing data should be handled. For example, when using MI, compatibility between the target analysis and imputation models is known to be important [11], but the backward induction nature of Q-learning means that correctly defining the imputation method is not straight forward.

In this paper, we aim to evaluate the performance of CCA and MI on the estimation of Q-learning parameters in a typical two-stage, two-treatment SMART. We begin by formalising several missing data mechanisms that might be found in a SMART context and reviewing Q-learning, before describing in detail the simulation study following the guidelines in Morris et al. [12]. We then present the results of the simulation study which compares the performance of CCA and MI for estimation of the stage 1 treatment effect under a range of missingness proportions, missing data mechanisms and varied stage 2 treatment effects. We also compare the different analytical approaches in a case study, using data from a smoking cessation SMART.

Methods

Missingness in SMART designs

A simple two-stage SMART design where randomisation at stage 2 is not dependent on the response to stage 1 is presented in Fig. 1. In this design we denote a single (binary) baseline characteristic by $O_1 = \{-1, 1\}$; treatments at stage 1 and stage 2 by $A_1 = \{-1, 1\}$ and $A_2 = \{-1, 1\}$ respectively; stage 1 responder status by $O_2 = \{-1, 1\}$; and the continuous outcome after stage 2 treatment by Y .

We begin by formalising four missingness scenarios that we might observe within a typical two-stage, two-treatment SMART design depicted using missing data directed acyclic graphs (m-DAGs, Fig. 2). The nodes M_{O_2} , M_{A_2} and M_Y in the m-DAG [6, 13] represent missingness in O_2 , A_2 and Y respectively. We assume that O_1 and A_1 are complete.

- Missing data scenario 1: Data records could be lost post data collection at both stages, which would lead to stage 1 responder status (O_2) and stage 2 outcome (Y) to be missing not dependent on any variables, but treatment at stage 2 (A_2) being fully observed.
- Missing data scenario 2: The final outcome (Y) could be missing dependent on the treatment given in stage 2 (A_2). For example, participants allocated to a less effective treatment arm could be more likely to drop out before the outcome data are collected. In addition

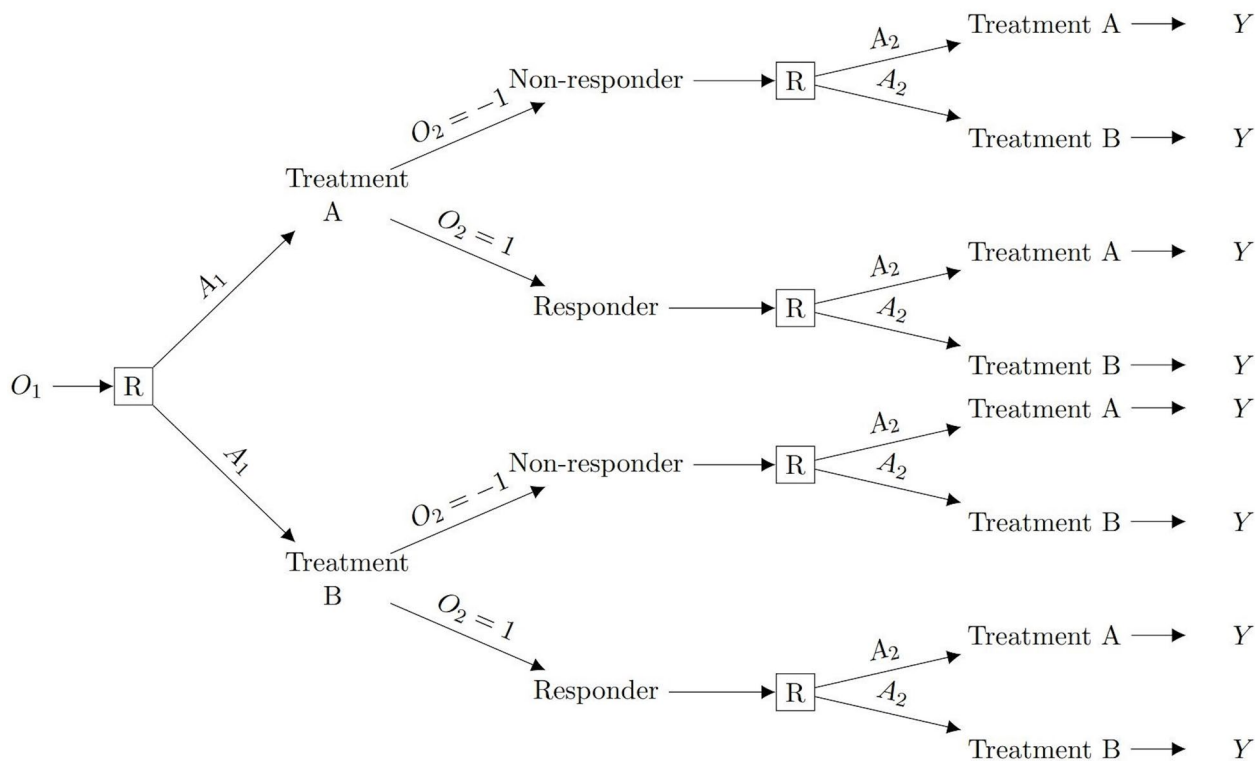


Fig. 1 Example of a SMART design where participants are first randomised in stage 1 and re-randomised in stage 2. Footnotes: ^a O_1 represents a baseline characteristic; A_1 and A_2 represent treatment at stage 1 and stage 2 respectively; O_2 represents stage 1 responder status; and Y represents the outcome after stage 2 treatment (also the final outcome). R is randomisation at each stage. ^b For simplicity, we present the tree diagram for those with $O_1 = 1$, this is repeated for those with $O_1 = -1$

to having missing data in the outcome, we could also have missingness in the stage 1 responder status (O_2) where (a) O_2 is missing not dependent on any variables; and (b) O_2 is missing due to a common cause between M_{O_2} and Y which induces an association.

- Missing data scenario 3: Participants could drop out after stage 1 leading to missing responder status in stage 1 (O_2), missing treatment assignment in stage 2 (A_2), and missing outcome data in stage 2 (Y). This could occur for many reasons including not being satisfied with the treatments, potential side effects or baseline characteristics (i.e. older participants being less likely to continue). Reasons for drop out could be related to (a) stage 1 treatment (A_1) and baseline characteristics (O_1) or (b) due to A_1 , O_1 and also a common cause between M_{O_2} and Y which induces an association.
- Missing data scenario 4: Participants who were responders to stage 1 treatment may be more likely to not agree to be randomised at stage 2 and therefore have missing treatment assignment (A_2) and outcome (Y) at stage 2. In this scenario, A_2 could be missing dependent on (a) O_2 or (b) O_2 and also a

common cause between M_{A_2} and Y which induces an association.

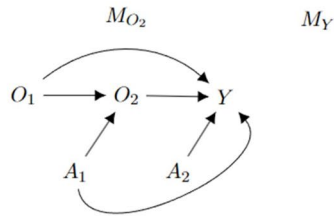
We chose to use m-DAGs to depict the missingness assumptions instead of the standard Rubin classification (MCAR/MAR/MNAR) because as stated by Lee et al. [6] the standard Rubin classifications are difficult to assess when there are missing data in multiple variables. For trials that consist of multiple stages such as SMARTs, there are often missing data in multiple variables (as depicted in our missing data scenarios described above) where a participant does not complete stage 1 and progress to stage 2 of the trial.

Q-learning

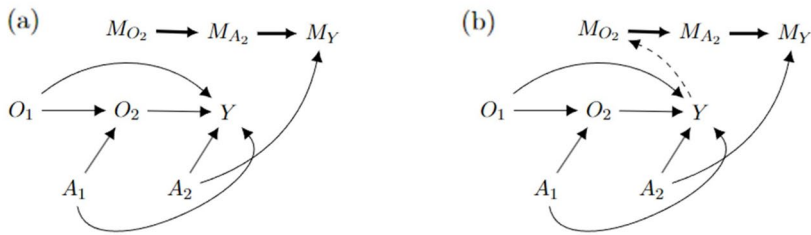
Q-learning is a stage-wise method where regression models are used to estimate the quality of each treatment decision at each stage using Q-functions, while accounting for variables from previous stages [5].

For a two-staged SMART, the stages are defined by k , where $k = 1, 2$. Here, O_k defines the variables observed before the k th stage; A_k is the treatment assigned at the k th stage where $A_k = \{-1, 1\}$; H_k is the history at the k th

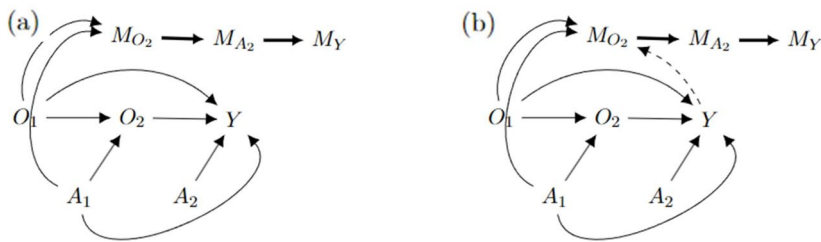
Missing data scenario 1



Missing data scenario 2



Missing data scenario 3



Missing data scenario 4

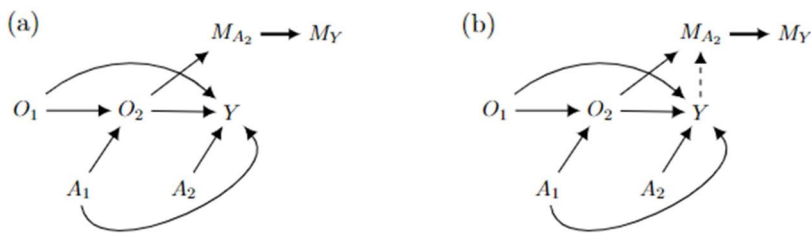


Fig. 2 m-DAGs depicting a two-stage SMART where the nodes M_z represent missingness in variable z . Footnotes: ^a Nodes M_{O_2} , M_{A_2} and M_Y represent missingness in O_2 , A_2 and Y respectively. ^b When O_2 is missing, both A_2 and Y are also missing as denoted by a thick black arrow. Graphs (b) are complements to (a) that include a dashed arrow representing missingness due to a common cause between Y and M_{O_2} or M_{A_2} which induces an association. ^c For missing data scenario 1) O_2 and Y are missing not dependent on any variables; 2a) O_2 is missing not dependent on other variables and Y missing dependent on A_2 ; 2b) O_2 missing dependent on Y and Y missing dependent on A_2 ; 3a) O_2 missing dependent on A_1 and O_1 ; 3b) O_2 missing dependent on A_1 , O_1 and Y ; 4a) A_2 missing dependent on O_2 ; 4b) A_2 missing dependent on O_2 and Y

stage where $H_1 = O_1$ and $H_2 = (O_1, A_1, O_2)$; and Y_k is the outcome at the end of the k th stage.

A DTR is a sequence of treatment decision rules. A two-stage DTR consists of two decisions (d_1, d_2) which depend on the history at that stage:

$$d_k(h_k) \in A_k, k = 1, 2 \tag{1}$$

The Q-functions, Q_k , represent the expected value of the outcome conditional on treatment and history for the two stages and are defined as follows [14]:

$$Q_2(H_2, A_2) = E[Y_2|H_2, A_2] \tag{2}$$

$$Q_1(H_1, A_1) = E \left[Y_1 + \max_{a_2} Q_2(H_2, a_2) | H_1, A_1 \right] \tag{3}$$

If the Q-functions defined above were known, then the optimal DTR for stage 1 and 2 would be:

$$d_k^{opt}(h_k) = \operatorname{argmax}_{a_k} Q_k(h_k, a_k), k = 1, 2 \tag{4}$$

However, in reality the true values of the Q-functions are unknown and therefore we use linear regression models to estimate the Q-functions:

$$Q_k(H_k, A_k; \beta_k, \psi_k) = \beta_k^T H_k + (\psi_k^T H_k) A_k, k = 1, 2, \tag{5}$$

where β_k and ψ_k are the regression coefficients that represent the effects of the history variables and the interaction between history variables and treatment, respectively.

For Q-learning, the optimal decision rule for the second stage is first identified and then subsequently incorporated into the first stage expectation, resulting in each Q-function being estimated via backward induction [2]. For example, for a two-stage design with a continuous primary outcome at each stage, linear regression is performed to determine the optimal decision rule at stage 2 first, based on the primary continuous outcome Y_2 and history (H_2, A_2) to obtain $(\hat{\beta}_2, \hat{\psi}_2)$:

$$\hat{d}_2^{opt}(h_2) = \operatorname{argmax}_{a_2} Q_2(h_2, A_2; \hat{\beta}_2, \hat{\psi}_2) = \operatorname{sign}(\hat{\psi}_2^T h_2) \tag{6}$$

Then we obtain an estimated outcome for stage 1 assuming that the optimal second stage outcome would be realised. This stage-1 outcome that combines data and expectations (often referred to as a pseudo-outcome) is defined as:

$$\begin{aligned} \hat{Y}_{1i} &= Y_{1i} + \max_{a_2} Q_2(H_2, a_2; \hat{\beta}_2, \hat{\psi}_2) \\ &= Y_{1i} + \hat{\beta}_2^T H_{2i} + |\hat{\psi}_2^T H_{2i}|, i = 1, \dots, n \end{aligned} \tag{7}$$

for the i th patient ($i = 1$ to n). Note that we can specify $Y_{1i} = 0$ if we only analyse one single primary outcome at the end of stage 2 (end of the study for both stages).

Regression is then performed at stage 1, based on the stage-1 pseudo-outcome \hat{Y}_{1i} and history (H_1, A_1) to obtain $(\hat{\beta}_1, \hat{\psi}_1)$, leading to the estimated first stage optimal decision rule, as defined below:

$$\hat{d}_1^{opt}(h_1) = \operatorname{argmax}_{a_1} Q_1(h_1, A_1; \hat{\beta}_1, \hat{\psi}_1) = \operatorname{sign}(\hat{\psi}_1^T h_1) \tag{8}$$

In a SMART the interest is typically in which treatment will give the best outcome at the end of the study, represented by the ψ parameters. For a two-staged trial the estimated optimal DTR is $(\hat{d}_1^{opt}, \hat{d}_2^{opt})$.

When we use Q-learning to analyse SMART data, an estimation issue, commonly referred to as ‘non-regularity’, can occur when at least some participants have no unique optimal treatment decision after stage 1, which can induce bias in the treatment effect estimation (see Additional file 2). In our simulation study described in the following section, we considered scenarios where this issue does and does not arise.

Simulation study

We used a simulation study to evaluate the performance of CCA and MI for handling missing data in a simple SMART to determine the optimal DTR using Q-learning.

Chakraborty et al. [14] proposed six SMART simulation settings which are either ‘regular’, ‘near-regular’ or ‘non-regular’ (see Additional file 2 for a description of these simulation settings). We considered the same simulation settings as it is known that non-regularity affects the estimation of certain Q-learning parameters (treatment effect parameters at any stage prior to the last). These simulation settings have since been used in several methodological studies [15–18]. We use three of Chakraborty’s simulation settings in our simulation study.

Data generation

We generated 1000 simulated datasets of $n = 500$ each using the data-generating models described below:

1. Participants were randomly assigned a treatment (binary) at stage 1 with the probability of 0.5.

$$P(A_1 = 1) = P(A_1 = -1) = 0.5 \tag{9}$$

2. A binary baseline characteristic O_1 , was generated with the probability of 0.5.

$$P(O_1 = 1) = P(O_1 = -1) = 0.5 \tag{10}$$

3. A binary stage 1 responder status O_2 was generated with the following probability, conditional on O_1 and A_1 :

$$\begin{aligned} P(O_2 = 1 | O_1, A_1) &= 1 - P(O_2 = -1 | O_1, A_1) \\ &= \operatorname{expit}(\delta_1 O_1 + \delta_2 A_1) \end{aligned} \tag{11}$$

where $\operatorname{expit}(x) = \exp(x)/(1 + \exp(x))$.

A participant is considered a responder if $O_2 = 1$ and a non-responder if $O_2 = -1$

- Participants' treatment allocation at stage 2 was generated as a binary variable with a probability of 0.5, with stage 1 treatment and responder status as stratification variables, that is, within each A_1 and O_2 stratum:

$$P(A_2 = 1) = P(A_2 = -1) = 0.5 \tag{12}$$

- Finally, a stage 2 continuous outcome was generated as $\mu_{Y2} = E(Y|O_1, A_1, O_2, A_2)$

$$\mu_Y = \gamma_1 + \gamma_2 O_1 + \gamma_3 A_1 + \gamma_4 O_1 A_1 + \gamma_5 A_2 + \gamma_6 O_2 A_2 + \gamma_7 A_1 A_2 + \varepsilon \tag{13}$$

where $\varepsilon \sim N(0, 1)$.

We considered 5 simulation settings that investigated a range of stage 1 and stage 2 treatment effects and represented different non-regularity settings, including simulation settings 1, 3 and 4 proposed by Chakraborty et al. [14] (corresponds to Treatment effect 1, 3 and 4 respectively) and two extra settings (Treatment effect 2 and 5) where stage 1 treatment effect did not equal 0:

- Treatment effect 1: (fully non-regular) there is no treatment effect on Y at either stage 1 or 2 for any participant.
- Treatment effect 2: (fully regular) there is a relatively large treatment effect on Y at stage 2 for every participant, and a treatment effect at stage 1 with a magnitude of -0.5 (see Table 1).
- Treatment effect 3: (non-regular) there is no treatment effect on Y at stage 2 for half of the participants, but a relatively large effect for the other half, and no treatment effect at stage 1.

- Treatment effect 4: (regular, but close to non-regular) there is a very weak treatment effect on Y at stage 2 for half of the participants, but a relatively large effect for the other half, and minimal treatment effect at stage 1.
- Treatment effect 5: (non-regular) there is no treatment effect on Y at stage 2 for half of the participants, but a relatively large effect for the other half, and a treatment effect at stage 1 with a magnitude of 0.5 (see Table 1). This simulation setting is similar to Treatment effect 3 but we assume that there is an even larger treatment effect at stage 2 for the other half.

We dropped simulation setting 2 (regular, but close to non-regular), setting 5 (non-regular) and setting 6 (fully regular) proposed by Chakraborty et al. [14] as they had a similar regularity setting to the settings we considered (Treatment effect 2, 3, 4 and 5). We added two extra simulation settings (Treatment effect 2 and 5) where there was a larger treatment effect at stage 1.

For the 5 simulation settings considered above, the different stage 2 treatment effects depended on their stage 1 responder status and stage 1 treatment assigned, and were achieved by modifying the parameter values used to generate the outcome data Y (Eq. 13). The parameter values are provided in Table 1.

Missingness

We set up 4 scenarios with different patterns and percentages of missing data as described below (Fig. 2). The parameter values used in these settings are described in Additional file 1: Table S1. All variables not mentioned in a scenario were considered fully observed.

Table 1 Parameter values used to generate the 5 different non-regularity settings

Treatment effect	Stage 1 responder status (O_2) ^a		Outcome (Y) ^b							Stage 1 treatment effect
	δ_1	δ_2	γ_1	γ_2	γ_3	γ_4	γ_5	γ_6	γ_7	ψ_{10}
1	0.5	0.5	0	0	0	0	0	0	0	0
2	0.5	0.5	0	0	-0.5	0	0	0	1	-0.5
3	0.5	0.5	0	0	-0.5	0	0.5	0	0.5	0
4	0.5	0.5	0	0	-0.5	0	0.5	0	0.49	-0.01
5	0.5	0.5	0	0	-0.5	0	1	0	1	0.5

^a Delta values used to define the probability of O_2

^b Outcome generated by Eq. 13: $\mu_Y = \gamma_1 + \gamma_2 O_1 + \gamma_3 A_1 + \gamma_4 O_1 A_1 + \gamma_5 A_2 + \gamma_6 O_2 A_2 + \gamma_7 A_1 A_2 + \varepsilon$ where $\varepsilon \sim N(0, 1)$

^c Degree of non-regularity is calculated by the probability of the linear combination $\gamma_5 + \gamma_6 O_2 + \gamma_7 A_1 = 0$ where the linear combinations tell us the stage 2 treatment effects for the different combinations of O_2 and A_1

^d ψ_{10} derived by combinations of the parameters of the generative models, including δ s used to generate the stage 1 responder status (O_2) and γ s used to generate the outcome

1. Missing data scenario 1: stage 1 responder status (O_2) and stage 2 outcome (Y) were set to missing not dependent on any variables.
2. Missing data scenario 2: stage 1 responder status (O_2) was set to missing not dependent on any variables. Stage 2 outcome data (Y) and treatment at stage 2 (A_2) were set to missing whenever there were missing data in O_2 . Additionally, stage 2 outcome data (Y) were set to missing dependent on A_2 as specified in the following logistic regression model:

$$\text{logit}[P(M_Y = 1)] = \alpha_0 + \alpha_1[A_2 = 1] \quad (14)$$

We also explored missing data scenario 2b, where we assumed that the missingness in O_2 was dependent on a common cause between M_{O_2} and Y which induces an association between M_{O_2} and Y , by generating the missingness in O_2 using the following logistic regression model:

$$\text{logit}[P(M_{O_2} = 1)] = \alpha_0 + \alpha_1 Y \quad (15)$$

3. Missing data scenario 3: stage 1 responder status (O_2) was set to missing using a logistic regression where the probability of missingness was dependent on baseline characteristics (O_1) and treatment at stage 1 (A_1):

$$\text{logit}[P(M_{O_2} = 1)] = \alpha_0 + \alpha_1[O_1 = 1] + \alpha_2[A_1 = 1] \quad (16)$$

In this scenario, stage 2 treatment (A_2) and the outcome (Y) were set to missing whenever there is missing data in stage 1 responder status (O_2).

We also considered missing data scenario 3b where the missingness in O_2 was additionally dependent on a common cause between M_{O_2} and Y which induces an association between M_{O_2} and Y . Therefore, the missingness in stage 1 responder status (O_2) was specified using the following logistic regression model:

$$\text{logit}[P(M_{O_2} = 1)] = \alpha_0 + \alpha_1[O_1 = 1] + \alpha_2[A_1 = 1] + \alpha_3 Y \quad (17)$$

4. Missing data scenario 4: if a participant was a responder to stage 1 treatment ($O_2 = 1$) then they were set to be more likely to have missing in stage 2 treatment (A_2) and outcome (Y). This was achieved using the following logistic regression model:

$$\text{logit}[P(M_{A_2} = 1)] = \alpha_0 + \alpha_1[O_2 = 1] \quad (18)$$

We also explored missing data scenario 4b where stage 2 treatment (A_2) was set to missing according to the following logistic regression model including the variables O_2 (stage 1 responder status) and Y (assuming a common cause of missingness in A_2 and Y):

$$\text{logit}[P(M_{A_2} = 1)] = \alpha_0 + \alpha_1[O_2 = 1] + \alpha_2 Y \quad (19)$$

The α parameters (except α_0) in the equation above were assigned based on the strength of association between the missing indicator and predictors of missingness, namely $\alpha = \log(\text{OddsRatio})$. For a weak association an odds ratio of 1.6 was used and for a strong association an odds ratio of 3 was used. The intercepts of the logistic regression models (α_0) above were chosen by iteration to achieve the required percentage of missingness (20% or 40%).

For each of the 5 simulation settings mentioned above which represented different stage 2 treatment effects, we considered the 7 missing data scenarios and for all missing data scenarios, except missing data scenario 1, we explored both weak and strong associations between the missing indicator and predictors of missingness. For all missing data scenarios, we also explored both 20% and 40% missingness. In total, we set up 130 simulation scenarios.

Estimand of interest and target analysis

Our interest lies in estimating the treatment effect at stage 1 estimated via the parameters of a Q-learning model at stage 1 from the two-stage, two-treatment SMART design.

These parameters were estimated using two-stage Q-learning process. First, we analysed outcome data at stage 2, using the following linear model to find the optimal treatment decision that maximises the Q_2 function:

$$Q_2(H_2, A_2) = \beta_{20} + \beta_{21}O_1 + \beta_{22}A_1 + \beta_{23}O_1A_1 + (\psi_{20} + \psi_{21}O_2 + \psi_{22}A_1)A_2 \quad (20)$$

Next, we obtained an estimated outcome for stage 1 (known as the stage 1 pseudo-outcome) for each individual, assuming that it is the average best outcome that can be obtained (Eq. 7).

Finally, we fitted a linear model to stage 1 data, based on the stage 1 pseudo-outcome, using the following model:

$$Q_1(H_1, A_1) = \beta_{10} + \beta_{11}O_1 + (\psi_{10} + \psi_{11}O_1)A_1 \quad (21)$$

The treatment effect at stage 1 (ψ_{10}), used to estimate the optimal decision rule in stage 1, is the estimand of interest.

We also explored the stage 2 treatment effect estimated via the parameters of the Q-learning model at stage 2 (ψ_{20}).

Methods to handle missing data

We compared the performance of MI and CCA for handling the missing data in each of the scenarios. For CCA, all participants with missing values in any of the variables were excluded from the analysis. For MI, the imputation model for Y (stage 2 outcome) included all the variables in the target analysis model, (including the interaction terms O_1A_1 , O_2A_2 and A_1A_2). To ensure compatibility with our analysis model, the imputation model for O_2 also included the interaction term YA_2 , and the imputation model for A_2 included the interaction terms YO_2 and YA_1 . Since the variables that need to be imputed consist of both binary and continuous variables, we used MI by chained equations where different univariate models can be used to impute the different types of variables with missing data [19]. We did not include any auxiliary variables in the imputation model and we return to this in the discussion.

According to the rule of thumb that the number of imputations should be at least equal to the proportion of missing data [20], for the scenarios where approximately 20% of the cases had missing data 20 imputations were generated, and for the scenarios where approximately 40% of the cases had missing data 40 imputations were generated.

Performance measures

Performance measures selected to compare the performance of CCA and MI were the absolute bias, the difference between the average estimate over 1000 datasets with the true value of ψ_{10} ; the empirical standard error, derived by square rooting the empirical variance; the mean squared error (MSE), derived by adding the variance and squared bias obtained above; bootstrap model-based standard errors, calculated by obtaining the standard error of the 200 bootstrap samples for each simulated dataset and then averaging these standard errors over the 1000 simulated datasets; and coverage, the proportion of 95% confidence intervals that contain the true value of ψ_{10} . Due to

long computational time, we selected treatment effect simulation settings 1,2 and 4 and for these selected two missing data scenarios (m-DAG 1 and m-DAG 2a) to present the bootstrap model-based standard errors and coverage. All data simulation and analyses were conducted using R version 4.3.1.

Real data example

To further compare the performance of the CCA and MI, we applied the two missing data methods to a real data example.

The real data example is based on two smoking cessation studies, namely the Project Quit (PQ) study [6] and the Forever Free (FF) study, which combined form a SMART. Both studies were conducted by the Center for Health Communications Research at the University of Michigan. The PQ study aimed to help participants quit smoking by using a web-based smoking cessation program and the FF study was a follow-on study to the PQ study where the purpose was to help participants of the PQ study who quit smoking to remain non-smokers and those who did not quit, quit smoking. Combining data from the two studies, we get a simple two-stage SMART where the PQ study represents the first stage, and the FF study represents the second stage of the SMART.

We analysed this dataset as a SMART using the baseline characteristics (O_1) gender, age, health maintenance organisation, education, number of cigarettes smoked, motivation and self-efficacy. Stage 1 intervention was whether the participant was randomised to a web-based smoking cessation program that did or did not include tailored hypothetical success stories. The binary variable collected at the end of stage 1 capturing whether a participant smoked in the last 7 days, was used as the stage 1 responder status (O_2). At stage 2 all participants (including both those who had quit smoking and those who had not quit smoking) were randomised to an intervention that requested them to read a specific version of the smoking cessation and relapse prevention booklets, wordings of these booklets were adjusted according to whether the participant had quit smoking or not. In the control group no booklets were handed out. The final outcome (Y) collected at the end of stage 2 was the number of attempts to quit smoking.

There were 1848 participants who were randomised to an intervention at stage 1 (PQ intervention). Only 479 participants who gave consent for the FF study were randomised to a treatment at stage 2. To have a dataset with a complete SMART design, our analysis was restricted to the 479 participants that received interventions at both stages. Therefore, by design, there were no participants

who dropped out of the study prior to being randomised at stage 2.

Results

Simulation Study

Figures 3, 4, 5 and 6 and Additional file 3: Figures S1–S6 summarise the performance of CCA and MI for estimating ψ_{10} , across the different simulation settings with varied stage 1 and 2 treatment effects and missing data scenarios described above.

Bias

For both CCA and MI, we observed close to zero bias under all missing data scenarios when there was no treatment effect at either stage 1 or 2 for any participants (Treatment effect 1, Fig. 3). When all participants had a relatively large treatment effect at stage 2 (Treatment effect 2, Fig. 4), we observed slightly more bias with MI than CCA under missing data scenarios 2b and 3a, albeit this bias was minimal. For missing data scenarios 2b and 3a, the bias was greater for stronger associations (\sim threefold increase compared to weaker relationships of \sim 1.6-fold increase) between missing indicators O_2 , A_2 and/or Y and their predictors of missingness.

When the treatment effect at stage 2 varied between participants (Treatment effect 3, 4, 5), for example when

half of the population had a very weak stage 2 treatment effect while the other half having a relatively large effect (Treatment effect 4, Fig. 5), MI showed greater bias than CCA. For these scenarios, we observed an increase in absolute bias for MI when the percentage of missing data increased from 20 to 40% but observed no change for CCA. For all simulation settings where treatment effect at stage 2 varied between participants, bias was not associated with the strength of associations between missing indicators and their predictors of missingness except for missing data scenario 2b (where O_2 missing dependent on Y and Y missing dependent on A_2) and 4b (where A_2 missing dependent on O_2 and Y) when half of the participants had no stage 2 treatment effect but the other half had a large stage 2 treatment effect (Treatment effect 5, Fig. 6), where there was greater bias for both missing data methods when the associations were stronger.

When estimating the stage 2 treatment effect, for both CCA and MI, we observed negligible bias for most missing data scenarios and simulation settings (Additional File 4: Tables S5-S9), except when the missingness in O_2 was dependent on a common cause between M_{O_2} and Y ; and missingness in the stage 2 outcome data was dependent on A_2 (m-DAG 2a, Fig. 2), where both methods showed bias.

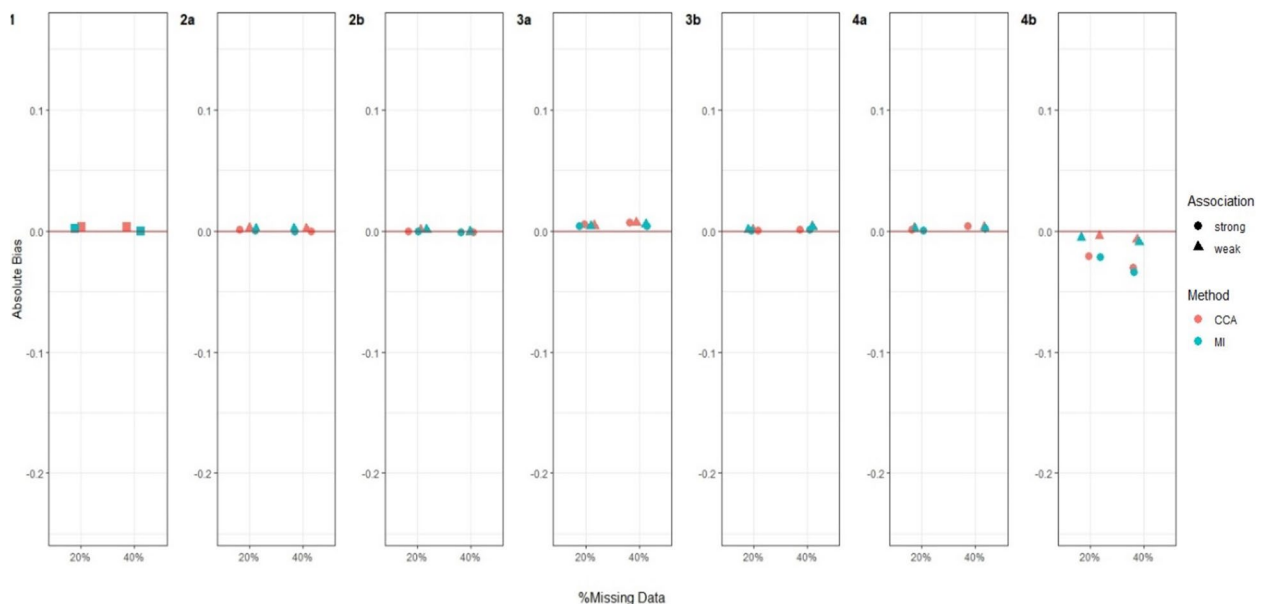


Fig. 3 Bias in estimating stage 1 treatment effect ψ_{10} under Treatment effect 1 simulation setting^a. Footnotes: ^a True value of $\psi_{10} = 0$ for Treatment effect 1 simulation setting when there is no treatment effect at either stage 1 or 2 for any participant. ^b Complete case analysis (CCA) and multiple imputation (MI) were used to handle missing data, where 20% or 40% had incomplete data under the seven missing data scenarios described in the Missingness in SMART designs section, see Fig. 2. ^c For weak association between the missing indicator and other variables (as described below) an OR of 1.6 was used; and for a strong association an OR of 3 was used. The other variables were: missing data scenario 2a) $A_2 \rightarrow M_Y$; 2b) $Y \rightarrow O_2$ and $A_2 \rightarrow M_Y$; 3a) $A_1 \rightarrow M_{O_2}$ and $O_1 \rightarrow M_{O_2}$; 3b) $A_1 \rightarrow M_{O_2}O_1 \rightarrow M_{O_2}$ and $Y \rightarrow M_{O_2}$; 4a) $O_2 \rightarrow M_{A_2}$; and 4b) $O_2 \rightarrow M_{A_2}$ and $Y \rightarrow M_{A_2}$. ^d Monte Carlo error less than 0.0035 for all estimates

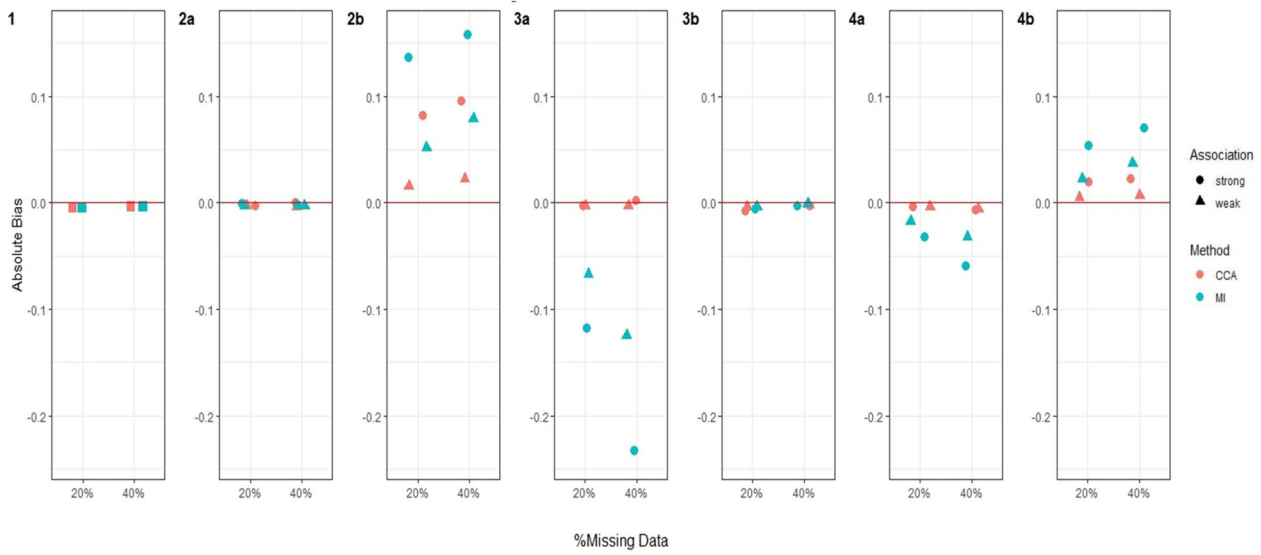


Fig. 4 Bias in estimating stage 1 treatment effect ψ_{10} under Treatment effect 2 simulation setting^a. Footnotes: ^aTrue value of $\psi_{10} = -0.5$ for Treatment effect 2 simulation setting when there is a relatively large treatment effect at stage 2 for every participant. ^b Complete case analysis (CCA) and multiple imputation (MI) were used to handle missing data, where 20% or 40% had incomplete data under the seven missing data scenarios described in the Missingness in SMART designs section, see Fig. 2. ^c For weak association between the missing indicator and other variables (as described below) an OR of 1.6 was used; and for a strong association an OR of 3 was used. The other variables were: missing data scenario 2a) $A_2 \rightarrow M_Y$; 2b) $Y \rightarrow M_{O_2}$ and $A_2 \rightarrow M_Y$; 3a) $A_1 \rightarrow M_{O_2}$ and $O_1 \rightarrow M_{O_2}$; 3b) $A_1 \rightarrow M_{O_2}O_1 \rightarrow M_{O_2}$ and $Y \rightarrow M_{O_2}$; 4a) $O_2 \rightarrow M_{A_2}$; and 4b) $O_2 \rightarrow M_{A_2}$ and $Y \rightarrow M_{A_2}$. ^d Monte Carlo error less than 0.0035 for all estimates

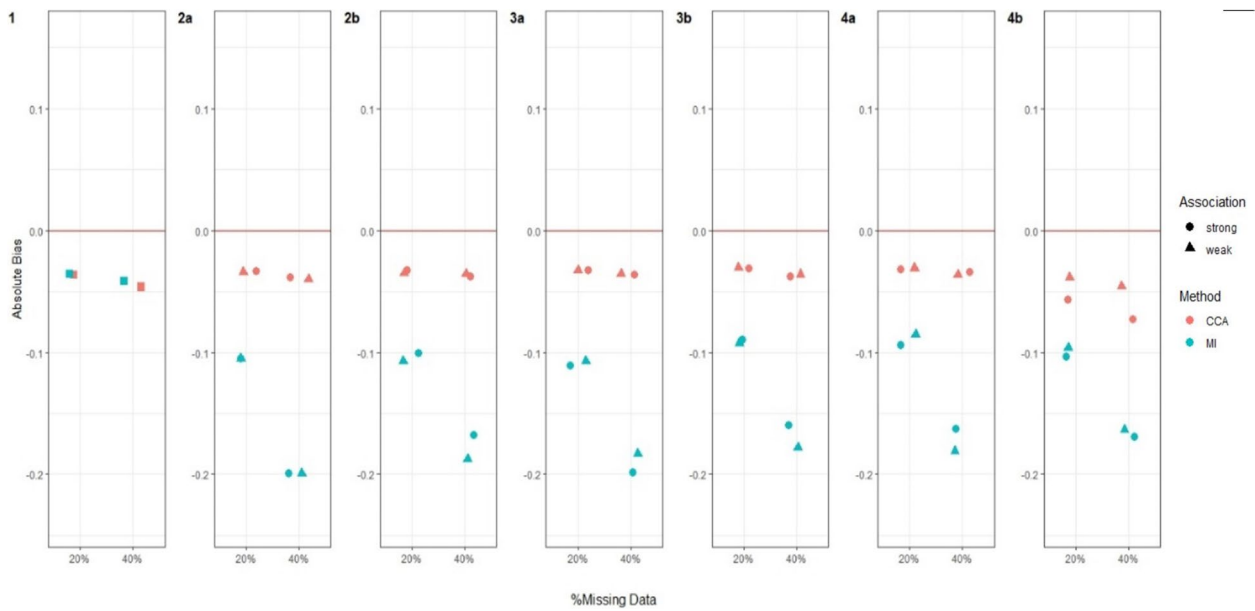


Fig. 5 Bias in estimating stage 1 treatment effect ψ_{10} under Treatment effect 4 simulation setting^a. Footnotes: ^aTrue value of $\psi_{10} = -0.01$ for Treatment effect 4 simulation setting when there is a very weak treatment effect at stage 2 for half of the participants, but a relatively large effect for the other half. ^b Complete case analysis (CCA) and multiple imputation (MI) were used to handle missing data, where 20% or 40% had incomplete data under the seven missing data scenarios described in the Missingness in SMART designs section, see Fig. 2. ^c For weak association between the missing indicator and other variables (as described below) an OR of 1.6 was used; and for a strong association an OR of 3 was used. The other variables were: missing data scenario 2a) $A_2 \rightarrow M_Y$; 2b) $Y \rightarrow M_{O_2}$ and $A_2 \rightarrow M_Y$; 3a) $A_1 \rightarrow M_{O_2}$ and $O_1 \rightarrow M_{O_2}$; 3b) $A_1 \rightarrow M_{O_2}O_1 \rightarrow M_{O_2}$ and $Y \rightarrow M_{O_2}$; 4a) $O_2 \rightarrow M_{A_2}$; and 4b) $O_2 \rightarrow M_{A_2}$ and $Y \rightarrow M_{A_2}$. ^d Monte Carlo error less than 0.0035 for all estimates

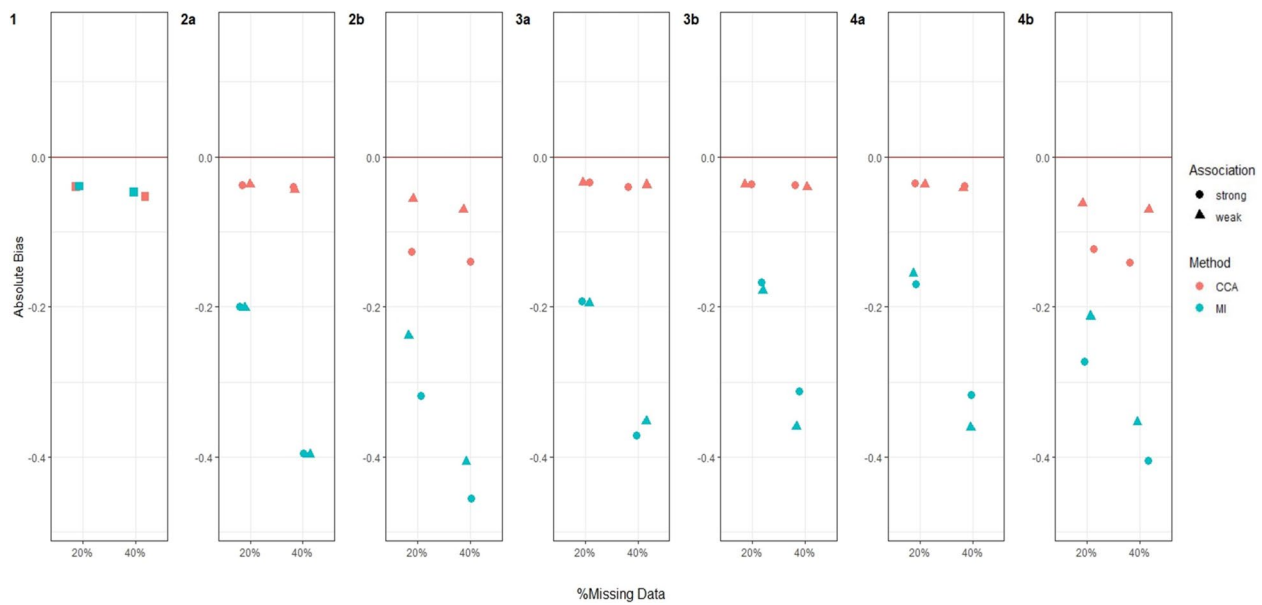


Fig. 6 Bias in estimating stage 1 treatment effect ψ_{10} under Treatment effect 5 simulation setting^a. Footnotes: ^a True value of $\psi_{10} = 0.5$ for Treatment effect 5 simulation setting when there is no treatment effect at stage 2 for half of the participants, but a relatively large effect for the other half. This simulation setting is similar to Treatment effect 3 but we assume that there is an even larger treatment effect at stage 2 for the other half. ^b Complete case analysis (CCA) and multiple imputation (MI) were used to handle missing data, where 20% or 40% had incomplete data under the seven missing data scenarios described in the Missingness in SMART designs section, see Fig. 2. ^c For weak association between the missing indicator and other variables (as described below) an OR of 1.6 was used; and for a strong association an OR of 3 was used. The other variables were: missing data scenario 2a) $A_2 \rightarrow M_Y$; 2b) $Y \rightarrow M_{O_2}$ and $A_2 \rightarrow M_Y$; 3a) $A_1 \rightarrow M_{O_2}$ and $O_1 \rightarrow M_{O_2}$; 3b) $A_1 \rightarrow M_{O_2}$, $O_1 \rightarrow M_{O_2}$ and $Y \rightarrow M_{O_2}$; 4a) $O_2 \rightarrow M_{A_2}$; and 4b) $O_2 \rightarrow M_{A_2}$ and $Y \rightarrow M_{A_2}$. ^d Monte Carlo error less than 0.0035 for all estimates

Empirical standard errors and Mean squared errors (MSE)

The empirical standard errors and MSE are shown in Additional file 3: Figures S2–S6. When there is a relatively large treatment effect at stage 2 for every participant (Treatment effect 2), we observed higher empirical standard errors for MI compared to CCA, under all missing data scenarios except for when stage 1 responder status and stage 2 outcome were missing not dependent on any variables (missing data scenario 1). For all other Treatment effect settings (e.g. no treatment effect at either stage 1 or 2, varying treatment effects at stage 2 between participants), the empirical standard errors were generally lower for MI compared to CCA under all missing data scenarios. As expected, empirical standard errors increased with the percentage of missingness. Stronger and weaker associations between the missing data indicators and their predictors of missingness showed similar empirical standard errors.

The MSE for CCA was similar across all treatment effect settings and missing data scenarios. When there was no treatment effect at either stage 1 or 2 for any participant, we observed similar MSE for both MI and

CCA under all missing data scenarios (Additional file 3: Figure S2). For all other treatment effect settings (Additional file 3: Figures S3–S6), under all missing data scenarios except for missing data scenario 1, MI showed greater MSE compared to CCA which increased with the proportion of missingness. For missing data scenario 1, we observed similar MSE for both MI and CCA.

The bootstrap model-based standard errors produced similar values to the empirical-based standard errors (Additional file 4: Tables S2–S4). Lower coverage was observed with the treatment effect 4 simulation setting (Additional file 4: Table S4).

Real data example

There were 479 participants that received interventions at both stages and 216 participants with outcome data available at stage 2 (55% missing in the outcome). For CCA, the point estimate of the stage 1 treatment effect, the estimated mean difference in number of attempts to quit smoking in participants using a web-based smoking cessation program compared to no program, was

-0.147 (95% confidence interval: -0.752, 0.338). In comparison, when MI was used to handle the missing data the estimate of the stage 1 treatment effect was -0.190 (95% confidence interval: -0.617, 0.244). Both CCA and MI provided similar estimates.

Discussion

In this simulation study, we evaluated the performance of CCA and MI for handling missing data when Q-learning is used to estimate the stage 1 treatment effect from the data of a typical two-stage SMART. We found that when stage 1 responder status and stage 2 outcome were missing not dependent on any variables (m-DAG 1, Fig. 2), we observed expected unbiased estimates for both CCA and MI across all simulation settings for the treatment effects. When there was no treatment effect at stage 1 or stage 2, both MI and CCA showed negligible absolute bias and similar empirical standard errors under all missing data scenarios. However, when the treatment effect at stage 2 varied between participants and data were missing dependent on other variables (for example, stage 1 responder status missing dependent on stage 1 treatment and baseline variables), MI showed greater bias which increased with the percentage missingness, while the bias for CCA remained minimal or negligible. When the treatment effect at stage 2 varied between participants, we observed lower or similar empirical standard errors for MI compared to CCA under all missing data scenarios. When all participants had a relatively large stage 2 treatment effect, we observed some, albeit minimal bias, with slightly greater bias for MI. The empirical standard errors were also higher for MI compared to CCA under all scenarios except for when data were missing not dependent on any variables. The stage 2 treatment effect was estimated via the parameters of the Q-learning model at stage 2 which is a standard linear regression model. We observed that under most missing data scenarios and simulation settings both CCA and MI showed negligible bias when estimating the stage 2 treatment effect.

In a case study of a smoking cessation study, we obtained similar estimates for both CCA and MI.

Many studies have shown that in standard, single stage trials, when outcome data are missing only dependent on observed data, MI shows an advantage over a CCA in handling missing data as it is able to capture the uncertainty in the missing data [9, 21]. However, for two stage trials such as SMARTs when the analysis is performed in stages using a backward induction approach (Q-learning), we found that MI induced bias while CCA showed negligible bias. When the treatment effect at stage 2 varied between participants, MI failed to capture these varied treatment effects. The backward induction nature of

Q-learning caused difficulty in selecting an appropriate imputation model for MI, especially in how to appropriately account for the pseudo-outcome derived in the first regression step and included in the second regression step. Compatibility of the target analysis and imputation model is known to be important when MI is used to handle missing data [11] and from our simulation study it appears likely that Q-learning causes issues with this compatibility. Therefore, these issues may have contributed to the bias seen in MI.

Only a single study has investigated methods to handle missing data in SMARTs previously. Shortreed et al. [10] proposed an MI strategy to overcome challenges to handle missing data in SMARTs and applied this strategy on data from a case study, the Clinical Antipsychotic Trials of Intervention and Effectiveness (CATIE) study. They used weighted regression to estimate the mean response for each treatment regime, where weights were used to adjust for the probability of randomisation (similar to inverse probability weighting (IPW) used to adjust for confounding). In Shortreed et al.'s paper, CCA produced lower and more variable estimates compared to MI. In our study, we evaluated how well MI performed when Q-learning was used to find the optimal DTR and looked at how the varying treatment effects at each stage affected the performance of CCA and MI. Our simulation results showed that CCA produced less biased estimates compared to MI. Despite these findings from simulation studies, when applying CCA and MI to real data examples (PQ study and FF study) both missing data methods showed similar results.

Our paper considered four missing data scenarios, depicted using m-DAGs [6], that we might observe in a simple two-stage two-treatment SMART. These m-DAGs allow us to clearly outline the causal relationships between variables and their missingness and are critical for complex settings such as SMARTs given the multiple stages. Using m-DAGs to understand the missing data assumptions which describe how the data become missing is important as they help us choose appropriate methods to handle the missing data. However, our results are restricted to the simulation scenarios proposed [14] and the Q-learning analysis method [5]. Further research is needed to emulate this study for different targets of analysis (e.g. the mean of embedded DTRs, calculating the mean outcome if all participants follow one of the DTRs), estimation methods (e.g. such as IPW, G-estimation and weighted regression [3]), different missing data methods (e.g. IPW) and missing data causal diagrams including auxiliary variables or when missingness in the outcome also depends on the outcome values itself. It would also be useful to conduct a theoretical evaluation of the analytic approaches considered here to further understand their performance in the context of SMARTs.

Conclusions

In conclusion, when handling missing data in a two-stage SMART, our results have shown that CCA is less biased compared to MI when estimating the stage 1 treatment effect. Since SMART designs have multiple stages, we recommend using m-DAGs to outline the missing data assumptions used in the analysis.

Abbreviations

CATIE	Clinical Antipsychotic Trials of Intervention and Effectiveness
CCA	Complete Case Analysis
DTR	Dynamic Treatment Regimen
FF	Forever Free
IPW	Inverse-probability Weighting
MI	Multiple Imputation
m-DAG	Missing Directed Acyclic Graph
MSE	Mean Squared Error
PQ	Project Quit
RCT	Randomised Controlled Trial
SMART	Sequential Multiple Assignment Randomised Trial

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12874-025-02595-1>.

Supplementary Material 1.
Supplementary Material 2.
Supplementary Material 3.
Supplementary Material 4.

Acknowledgements

Many thanks to Principal Investigator (Victor J. Strecher, University of Michigan, US) and Bibhas Chakraborty (Duke NUS, Singapore) for making the data available.

Authors' contributions

RM, JAS and JX conceived of and designed the simulation study with assistance from ADS and KJL. JX drafted the manuscript with input from all co-authors. All authors were responsible for critical revision of the manuscript and have read and approved the final version.

Funding

JAS and KJL acknowledge support from a National Health and Medical Research Council (NHMRC) Investigator Grant Leadership Level 1 (IDs 1196068 and 2017498 respectively), and JX was partially supported by a NHMRC Centre of Research Excellence grant (ID 1171422), the Australian Trials Methodology (AusTriM) Research Network. Design of the simulation study, analysis of the case study, interpretation of data, and writing of the manuscript were done completely independently of any funding bodies.

Data availability

The statistical computing code for this simulation study is available on GitHub https://github.com/jessicaxu3205/SMART_missing. Data from the Project Quit (PQ) study and the Forever Free (FF) study were made available by the Principal Investigator (Victor J. Strecher, University of Michigan, US) and Bibhas Chakraborty (Duke NUS, Singapore) and are available from the authors upon reasonable request and with permission of Victor J. Strecher and Bibhas Chakraborty.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

¹Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, University of Melbourne, Melbourne, VIC, Australia. ²Clinical Epidemiology and Biostatistics Unit, Murdoch Childrens Research Institute, Royal Children's Hospital, Melbourne, VIC, Australia. ³Department of Paediatrics, University of Melbourne, Melbourne, VIC, Australia. ⁴Nuffield Department of Medicine, University of Oxford, Oxford, UK.

Received: 12 May 2024 Accepted: 13 May 2025

Published online: 01 July 2025

References

1. Wahed AS, Tsiatis AA. Optimal estimator for the survival distribution and related quantities for treatment policies in two-stage randomization designs in clinical trials. *Biometrics*. 2004;60(1):124–33.
2. Liu Y, Zeng D, Wang Y. Use of personalized Dynamic Treatment Regimes (DTRs) and Sequential Multiple Assignment Randomized Trials (SMARTs) in mental health studies. *Shanghai Arch Psychiatry*. 2014;26(6):376–83.
3. Mahar RK, McGuinness MB, Chakraborty B, Carlin JB, MJ JI, Simpson JA. A scoping review of studies using observational data to optimise dynamic treatment regimens. *BMC Med Res Methodol*. 2021;21(1):39.
4. Murphy SA. An experimental design for the development of adaptive treatment strategies. *Stat Med*. 2005;24(10):1455–81.
5. Nahum-Shani I, Qian M, Almirall D, Pelham WE, Gnagy B, Fabiano GA, et al. Q-learning: a data analysis method for constructing adaptive interventions. *Psychol Methods*. 2012;17(4):478–94.
6. Lee KJ, Carlin JB, Simpson JA, Moreno-Betancur M. Assumptions and analysis planning in studies with missing data in multiple variables: moving beyond the MCAR/MAR/MNAR classification. *Int J Epidemiol*. 2023;52(4):1268–75.
7. Sterne JA, White IR, Carlin JB, et al. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. *BMJ*. 2009;338:b2393.
8. Lee KJ, Simpson JA. Introduction to multiple imputation for dealing with missing data. *Respirology*. 2014;19(2):162–7.
9. Sullivan TR, White IR, Salter AB, Ryan P, Lee KJ. Should multiple imputation be the method of choice for handling missing data in randomized trials? *Stat Methods Med Res*. 2016;27(9):2610–26.
10. Shortreed SM, Laber E, Scott Stroup T, Pineau J, Murphy SA. A multiple imputation strategy for sequential multiple assignment randomized trials. *Stat Med*. 2014;33(24):4202–14.
11. Bartlett JW, Seaman SR, White IR, Carpenter JR; Alzheimer's Disease Neuroimaging Initiative*. Multiple imputation of covariates by fully conditional specification: Accommodating the substantive model. *Stat Methods Med Res*. 2015;24(4):462–487.
12. Morris TP, White IR, Crowther MJ. Using simulation studies to evaluate statistical methods. *Stat Med*. 2019;38(11):2074–102.
13. Moreno-Betancur M, Lee KJ, Leacy FP, White IR, Simpson JA, Carlin JB. Canonical Causal Diagrams to Guide the Treatment of Missing Data in Epidemiologic Studies. *Am J Epidemiol*. 2018;187(12):2705–15.
14. Chakraborty B, Murphy S, Strecher V. Inference for non-regular parameters in optimal dynamic treatment regimes. *Stat Methods Med Res*. 2010;19(3):317–43.
15. Moodie EE, Chakraborty B, Kramer MS. Q-learning for estimating optimal dynamic treatment rules from observational data. *Can J Stat*. 2012;40(4):629–45.
16. Chakraborty B, Laber EB, Zhao Y. Inference for optimal dynamic treatment regimes using an adaptive m-out-of-n bootstrap scheme. *Biometrics*. 2013;69(3):714–23.
17. Simoneau G, Moodie EEM, Platt RW, Chakraborty B. Non-regular inference for dynamic weighted ordinary least squares: understanding the impact of solid food intake in infancy on childhood weight. *Biostatistics*. 2018;19(2):233–46.
18. Murray TA, Yuan Y, Thall PF. A Bayesian Machine Learning Approach for Optimizing Dynamic Treatment Regimes. *J Am Stat Assoc*. 2018;113(523):1255–67.

19. Azur MJ, Stuart EA, Frangakis C, Leaf PJ. Multiple imputation by chained equations: what is it and how does it work? *Int J Methods Psychiatr Res.* 2011;20(1):40–9.
20. White IR, Royston P, Wood AM. Multiple imputation using chained equations: Issues and guidance for practice. *Stat Med.* 2011;30(4):377–99.
21. Groenwold RH, Donders AR, Roes KC, Harrell FE Jr, Moons KG. Dealing with missing outcome data in randomized trials and observational studies. *Am J Epidemiol.* 2012;175(3):210–7.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.