Article:

Alshreef, A. orcid.org/0000-0003-2737-1365, Latimer, N. orcid.org/0000-0001-5304-5585, Tappenden, P. et al. (1 more author) (2025) Assessing methods for adjusting estimates of treatment effectiveness for patient nonadherence in the context of time-to-event outcomes and health technology assessment: a simulation study. Medical Decision Making, 45 (1). pp. 60-73. ISSN 0272-989X

https://doi.org/10.1177/0272989x241293414

Title: Assessing Methods for Adjusting Estimates of Treatment Effectiveness for Patient Non-Adherence in the Context of Time-to-event Outcomes and Health Technology Assessment: A Simulation Study

Running head: Assessing methods for adjusting for non-adherence

Authors and affiliations:

Abualbishr Alshreef, PhD[1], Nicholas Latimer, PhD[2,3], Paul Tappenden, PhD[2], Simon Dixon, PhD[2]

[1] AbbVie, Inc. North Chicago, USA

[2] Health Economics and Decision Science, School of Health and Related Research (ScHARR), University of Sheffield, Sheffield, UK

[3] Delta Hat Ltd, Nottingham, UK

Corresponding Author:

Name: Abualbishr Alshreef

Address: AbbVie, Inc. 1 N Waukegan Rd, North Chicago, IL 60064, USA

Tel: +1 (630) 263-6313

Email: abualbishr.alshreef@abbvie.com or abu.alshreef@gmail.com

Co-authors contact details

NL, PT, SD: Health Economics and Decision Science, School of Health and Related Research, University of Sheffield, Regent Court, 30 Regent Street, Sheffield, S1 4DA, UK

Acknowledgements

This work was done in ScHARR, University of Sheffield, UK. An abstract from this work was presented as an oral presentation at the 43rd Annual Meeting of the Society for Medical Decision Making (SMDM) in Washington D.C., USA (October 19, 2021).

This publication is independent research funded by a National Institute for Health Research (NIHR) Doctoral Research Fellowship awarded to Dr Abualbishr Alshreef (DRF-2017-10-025). The views expressed in this publication are those of the authors and not necessarily those of the NHS, the NIHR or the UK Department of Health and Social Care.

Financial support for this study was provided entirely by a grant from the National Institute for Health Research (NIHR). The funding agreement ensured the authors' independence in designing the study, interpreting the data, writing, and publishing the report.

Acknowledgements

This report is independent research funded by a National Institute for Health Research (NIHR) Doctoral Research Fellowship awarded to Dr Abualbishr Alshreef (DRF-2017-10-025). The views expressed in this report are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health and Social Care.

We thank Ian White, PhD; Michael Crowther, PhD; William McKane, PhD, FRCP; and James Fotheringham, PhD, MRCP for their advice and/or comments on the simulation study protocol.

Conflict of Interests

AA works for AbbVie, Inc. and reported equity ownership in a publicly traded company (AbbVie, Inc., North Chicago, IL).
NL is supported by Yorkshire Cancer Research (Award reference number S406NL).
PT and SD declare that they have no conflict of interest.

Published version

This paper was published on 8th November 2024 by Medical Decision Making. This is the accepted version. The published version is available from: https://doi.org/10.1177/0272989X241293414
Alshreef A, Latimer N, Tappenden P, Dixon S. Assessing Methods for Adjusting Estimates of Treatment Effectiveness for Patient Non-Adherence in the Context of Time-to-event Outcomes and Health Technology Assessment: A Simulation Study. Medical Decision Making. First published online November 8, 2024. DOI: 10.1177/0272989X241293414

ABSTRACT

Purpose. We aim to assess the performance of methods for adjusting estimates of treatment effectiveness for patient non-adherence in the context of health technology assessment (HTA) using simulation methods.

Methods. We simulated trial datasets with non-adherence, prognostic characteristics, and a time-to-event outcome. The simulated scenarios were based upon a trial investigating immunosuppressive treatments for improving graft survival in patients who had had a kidney transplant. The primary estimand was the difference in restricted mean survival times in all patients had there been no non-adherence. We compared generalized methods (g-methods; marginal structural model with inverse probability of censoring weighting [IPCW], structural nested failure time model [SNFTM] with g-estimation) and simple methods (intention-to-treat [ITT] analysis, per-protocol [PP] analysis) in 90 scenarios each with 1900 simulations. Methods performance was primarily assessed according to bias.

Results. In implementation non-adherence scenarios, the average percent bias was 20% (ranging from 7 to 37%) for IPCW, 20% (8-38%) for SNFTM, 20% (8-38%) for PP, but 40% (20-75%) for ITT. In persistence non-adherence scenarios, average percent bias was 26% (9-36%) for IPCW, 26% (14-39%) for SNFTM, 26% (14-36%) for PP, but 47% (16-72%) for ITT. In initiation non-adherence scenarios, percent bias ranged from -29 to 110% for IPCW, -34 to 108% for SNFTM, -32 to 102% for PP, but between -18 to 200% for ITT.

Conclusion. In this study, g-methods and PP produced more accurate estimates of the treatment effect adjusted for non-adherence than the ITT analysis. However, considerable bias remained in some scenarios.

HIGHLIGHTS

- Randomised controlled trials are usually analysed using the intention-to-treat (ITT) principle, which produces a valid estimate of effectiveness relating to the underlying trial, but when patient adherence to medications in the real world is known to differ from that observed in the trial, such estimates are likely to result in a biased representation of real-world effectiveness and cost-effectiveness.
- Our simulation study demonstrates that generalized methods (g-methods; IPCW, SNFTM) and per-protocol analysis provide more accurate estimates of the treatment effect than the ITT analysis, when adjustment for non-adherence is required; however, even with these adjustment methods, considerable bias may remain in some scenarios.
- When real-world adherence is expected to differ from adherence observed in a trial, adjustment methods should be used to provide estimates of real-world effectiveness.

INTRODUCTION

Patient non-adherence to medications can have a significant negative impact on the clinical effectiveness of treatments, which has a direct impact on cost-effectiveness. Randomised controlled trials (RCTs) are usually analysed using the intention-to-treat (ITT) principle, which produces a valid estimate of effectiveness relating to the underlying trial, but when real-world adherence is known to differ from that in the trial, such estimates are likely to result in a biased representation of real-world effectiveness. Efficacy estimates from RCTs may differ from real-world effectiveness due to their stringent selection criteria and other intercurrent events that might affect treatment effect estimates. One feature of these is non-adherence, and we aim to focus specifically on this in this paper. For instance, in a case study of statins, adherence levels based on trial evidence were estimated to range from 81% to 99%, whereas in pragmatic studies in community settings, adherence ranged from 21% to 87%.[1] If adherence in standard clinical practice differs from that observed in RCTs, it is likely to be necessary to use adjusted analyses to estimate the effectiveness (and cost-effectiveness) expected in the real world, departing from ITT analyses.[2-5] This is particularly important in the context of health technology assessment (HTA), as decisions are made for the real-world population.[6]

Adherence to medication is defined as *'the process by which patients take their medications as prescribed'*.[7] An influential taxonomy of adherence to medications set out by Vrijens et al.[7] identifies three components: (a) initiation of the treatment (when the first dose is taken by the patient); (b) implementation of the dosing regimen (to what extent the actual dosage of a patient corresponds to the prescribed dosing regimen); and (c) discontinuation (end of therapy).[7] Based on this taxonomy, non-adherence can occur in one or a combination of three situations: late or non-initiation, sub-optimal implementation, and/or early discontinuation (non-persistence).[7, 8] A number of other terms are commonly used in the literature to describe non-adherence including 'non-compliance' and 'non-concordance'.[9] Whilst some authors suggest differences between the precise meanings of these different terms, for the purpose of this paper, we treat them as being synonymous with non-adherence, and the Vrijens et al. taxonomy is used as a standard.

A previous review identified 12 methods for adjusting estimates of treatment effectiveness for non-adherence in the context of survival analysis and HTA.[6] Generalized methods (g-methods; marginal structural model with inverse probability of censoring weighting [IPCW], structural nested failure time model [SNFTM] with g-estimation, rank-preserving structural failure time model [RPSFTM] with g-estimation) and pharmacometrics-based methods using pharmacokinetics and pharmacodynamics (PKPD) analysis were found to be the most suitable for adjusting for non-adherence.[6] However, further research was warranted to assess the performance of these methods in simulation studies.[6]

The aim of this article is to assess the relative performance of non-adherence adjustment methods using a simulation study to provide recommendations for application in an HTA context.

METHODS

Simulation Study Design

The simulation study design is based on recommended methods.[10 11] A pre-specified study protocol was developed using the ADEMP (aims, data-generating mechanisms, estimands, methods, and performance measures) structural approach for planning simulation studies; this is summarised in the subsequent subsections.[11]

Aim

The study aimed to assess the performance of alternative non-adherence adjustment methods in estimating the treatment effect across a range of 90 scenarios using simulated RCT datasets with a time-to-event outcome, in which non-adherence was simulated to occur. The simulation was based on evaluating the treatment effect of hypothetical maintenance immunosuppressive drugs on graft survival for adult kidney transplant patients. Although the simulation was focused on kidney transplantation, the findings are likely to be applicable to other disease areas in studies with a time-to-event outcome where non-adherence is identified as an issue.

Data-Generating Mechanisms

Complex data-generating mechanisms (DGMs) were used to allow the assessment of alternative non-adherence adjustment methods across a range of pre-specified scenarios. These involve the specification of the pattern of non-adherence, prognostic variables, distributions, and covariate correlation structures. The scenarios covered alternative representations of seven factors thought to influence estimated efficacy and its uncertainty. The factors are: type of non-adherence (initiation, implementation, or persistence, defined below); level of non-adherence; sample size; the pattern of hazards (as represented by an underlying graft survival model); treatment effect size; relationship between treatment effect and adherence level; and the existence of any time-dependent treatment effect. A binary variable was used for simulating each type of non-adherence, with the probabilities of non-adherence at each time interval simulated such that the overall non-adherence patterns were classified as high/low. These were numerically defined as relative values depending on the type of non-adherence (e.g. 10% low implementation non-adherence and 40% high implementation non-adherence). The probability of non-adherence in the control arm was different to the experimental arm. This mimics the usual pattern of non-adherence seen in clinical trials. The values for the probability of non-adherence were assumed with values varied depending on follow-up time points and the type of non-adherence evaluated in each set of scenarios (more details are provided in Supplementary Appendix A).

By varying these factors across scenarios, we were left with 90 scenarios (see Table 1). The choice of these factors and the levels chosen for each are described in the following subsections. A parametric simulation approach was used to compare the performance of each method against a known "truth", as specified in the simulation program subsection below. In our simulation study, the truth represents what would be observed with zero non-adherence.

A directed-acyclic graph (DAG) presented in Figure 1 shows the relationships between covariates incorporated in the data-generating model, including randomisation, baseline and time-varying covariates, non-adherence, and graft survival outcome. The DAG was drawn based on evidence from the literature and discussion with two clinical experts (WM and JF). The DAG was used to conceptualise and guide the process of simulating the datasets.

The simulation assumed no "non-administrative" censoring due to loss of follow-up and no missing data. This assumption allowed the simulation to focus on addressing the issue of non-adherence without the need to simultaneously address other inter-current events.

Patterns of Medication Non-Adherence and Confounding Factors

Three types of non-adherence were included in this simulation study using binary variables: late or non-initiation (when the first dose is taken by the patient), sub-optimal implementation (to what extent the actual dosage of a patient corresponds to the prescribed dosing regimen), and/or early discontinuation (end of therapy or non-persistence).[7] Initiation and persistence were simulated using a binary variable that reflects the non-adherence event at a particular time point, whilst implementation reflects non-adherence during a particular time interval (e.g., 4 to 8 months). Patients non-adhering at the previous time point were modelled to remain non-adherent for the rest of the study follow-up.

A range of potential confounding factors was considered including patient characteristics which may predict non-adherence patterns, biomarkers and prognostic factors associated with non-adherence, and graft loss. A subset of the factors were incorporated as covariates in the simulated datasets based on discussion with two clinical experts and consideration of published literature.[12]

These were age as a baseline covariate and body mass index (BMI) as a time-dependent confounder.

Simulating Baseline Covariates and Randomisation Assignment

We simulated datasets for a two-arm RCT with 1:1 random allocation with 12 months of follow up. Data generation started with specifying the number of observations (sample size) followed by creating a baseline covariate (age), followed by the values of the time-dependent covariate (BMI) at baseline. BMI was assumed to be independent of age. This is followed by assigning observations to the experimental group or control group using *"randomize"*, a user-written command in Stata for performing the randomisation procedure.[13]

Simulating Non-Adherence and Time-Dependent Covariates

Each type of non-adherence was simulated for three-time intervals (baseline to 4 months, 4 - 8 months, and 8 – 12 months). Time-dependent covariates (i.e., BMI) were measured at baseline and two follow-up time points (4 and 8 months). As shown in the DAG (Figure 1), time-varying covariates (L's) and non-adherence variables (A's) were generated sequentially. For each follow-up time point, the non-adherence variable by treatment group was simulated first. This was dependent on the history of time-varying covariates and non-adherence at the previous time point.

In the simulated datasets, immunosuppressive treatment has a direct causal effect on graft survival; however, this causal relationship is confounded by baseline age and BMI measured at each follow-up time point. This makes BMI a time-dependent confounder (where the variable's value is influenced by previous treatment/adherence, influences subsequent treatment/adherence, and is a predictor of graft survival – see Figure 1).

Graft Survival Time Data-Generating Methods

Simulation of survival data was based on parametric survival models (PSM).[14] The generated datasets were checked to ensure their resemblance to realistic situations[15] before using the simulated datasets for assessing non-adherence adjustment methods. This was achieved by using summary statistics, Kaplan-Meier survival curves and model fit statistics using the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC).

Two graft survival-time data-generating models were used to simulate the RCT datasets: (i) a PSM with a Weibull distribution, and (ii) a two-component PSM with a mixture Weibull-Weibull distribution. The two models were chosen to improve transferability beyond the context of kidney transplantation whereby the shape of the survival curves differ by disease area and intervention.

To generate the truth (defined earlier) for each scenario, graft survival time outcomes were simulated using 1 million iterations for each scenario. This was implemented using the *survsim* Stata command by incorporating all baseline and updated time-varying covariates at all time-points except with perfect adherence.

Parameter Values and Distributions

The seven simulation factors (defined earlier) were varied in a partial factorial design to specify a range of realistic scenarios. The parameter values for the levels of non-adherence, the correlation between non-adherence and outcome and the other factors (defined by the DGMs) were varied to test the methods in a range of scenarios.

The partial factorial design resulted in 90 scenarios across the three types of non-adherence (Initiation= 38, implementation=34, persistence=18). More scenarios were specified for initiation and implementation non-adherence as these are considered the most important types for chronic diseases. The levels specified for each factor are shown in Table 1. The parameter values specified for each factor that were used to simulate the dataset are provided in Supplementary Appendix A.

Simulation Program

The *"simulate"* approach in Stata was used to perform a Monte Carlo simulation. The simulation program was run for a total of 90 scenarios for applying non-adherence adjustment methods with 1900 simulations each. The technical details and code used to simulate the datasets are provided in Supplementary Appendix C.

Estimands

The primary estimand of interest is the treatment effect using the difference in restricted mean graft survival times [RMST][16] between treatment groups, had there been no non-adherence. We also recorded experimental group RMST, control group RMST, and the population-level hazard ratio (HR) as estimates for secondary estimands, but these were not used for assessing methods performance. The calculation of RMST and HR involves fitting a flexible parametric model (FPM) to the graft survival data estimated by each method under the hypothetical assumption of zero non-adherence, restricted to the 1-year end-of-follow-up timepoint.

Methods Assessed

Each simulated RCT dataset was analysed using each of the following non-adherence adjustment methods. The methods assessed are described in more detail in Alshreef et al., 2019[6] including the estimands, estimators, and key assumptions. The application of each method in this simulation study is briefly described in the following subsections.

Intention-To-Treat Analysis

The ITT analysis involves fitting an FPM to both treatment groups combined (i.e., with treatment group as a covariate) using the *"stpm2"* user-written command in Stata. The FPM model was adjusted for age as a baseline covariate using 2 degrees of freedom in the model specification. The *"standsurv"* post estimation command was used to produce the difference in RMSTs between treatment groups alongside the standard error and confidence intervals around the estimates. RCTs are usually analysed using ITT analyses, which provide an unbiased comparison of the outcomes observed in the randomised groups. However, if patient adherence to medications in the real world differs from that observed in the trial, such ITT estimates are likely to result in a biased representation of real-world effectiveness which is an important consideration in the HTA context. The ITT method does not make any adjustment for non-adherence, but it was important to include this to demonstrate how biased estimates of the treatment effect based on an ITT analysis can be, if our objective is to obtain an estimate of the treatment effect adjusted for non-adherence. Thus, it is important to include ITT analysis as a benchmark in this simulation study.

Per-Protocol Analysis

The per-protocol (PP) analysis was implemented by censoring all non-adherent patients at the first time of non-adherence. Then the *stpm2* and *standsurv* commands were applied in the same way as described in the ITT analysis.

Marginal Structural Model with Inverse Probability of Censoring Weighting

The application of IPCW starts with creating a time-dependent non-adherence indicator for each time interval within the dataset (i.e., 0-4 months, 4-8 months, and 8-12 months). Then, a time-dependent outcome for graft loss was created using the same time intervals. To derive stabilised weights, four non-adherence logistic models were fitted (two models per treatment arm).

In these non-adherence models, we used logistic regression to predict non-adherence given baseline covariates and time in the control arm (Non-adherence Model 1). Non-adherence Model 2 was then fitted on the control group with both baseline and time-dependent covariates with interaction terms included in the model specification. Non-adherence Models 3 and 4 were fitted in the same way to the experimental group. Following the application of each logistic regression (non-adherence models), the "*predict*" command was used to estimate the probability of non-adherence for each patient observation included in the regression. These were then used to calculate the probability of remaining uncensored (i.e., the probability of adherence) by subtracting the probability of non-adherence from 1. This was undertaken at the individual patient-level using the estimates obtained from each of the non-adherence models.

To generate stabilised weights for the control group, we divided the probability of remaining uncensored (adherent) obtained from Model 1 by the probability of adherence generated from Model 2. The same was done for the experimental group using Models 3 and 4.

A pseudo-population (i.e., weighted) dataset was created using the stabilised IPCW weights and with non-adherers censored at the point that they first non-adhered – representing a population in which there was zero non-adherence. Then, we applied the *stpm2* model (including the *standsurv* post-estimation command) as described above to obtain the required RMST estimates. The model used robust standard errors to get 95% confidence intervals (CIs) that take into account the weighting of the data and the clustering of individuals.

Structural Nested Failure Time Model with G-estimation

The application of the SNFTM with g-estimation method was implemented as follows. First, we specified baseline and time-varying confounders for inclusion in the g-estimation model. This involved specifying age as a baseline covariate and BMI as a time-dependent confounder. The time-to-event indicator (time-dependent graft loss) was made time-dependent in the dataset. Then, time-lagged non-adherence variables were generated using individual patient-level adherence data for each time interval.

Second, we estimated the Acceleration Factor (AF) as the effect of time-dependent non-adherence on the graft survival time outcome. This was done for each treatment arm separately using the *"stgest3"* command, which is a Stata program for implementing g-estimation in an SNFTM.[17] The *stgest3* model incorporated the interaction of confounders with time in the same way applied to the IPCW weighting models.

Third, we used the estimated AF to adjust graft survival times by treatment arm allowing for recensoring. This results in estimated graft survival times that would have been observed if there had been no non-adherence.

Finally, the *stpm2* and *standsurv models* were used for generating the required estimates, as explained in previous sections. The CIs obtained do not factor in the fact the data have been adjusted, and therefore, are likely to underestimate uncertainty, with potential implications for coverage.

Performance Measures

Performance assessment was focused on the following key properties of the estimators: bias, accuracy, coverage, and both empirical and model-based standard errors. In addition, to quantify simulation uncertainty over the number of simulations, we estimated the Monte Carlo standard errors (MCSE) of the estimated performance measures. The formulae used to compute the performance measures are provided in Supplementary Appendix A. Convergence was also captured – for some models, such as IPCW, multiple models needed to converge in order to achieve successful estimation for one simulation. The results of the simulations were analysed using the *simsum* Stata command to generate the performance results.[18]

Simulation Summary

The simulation process and outputs are summarised in Box 1. The 90 simulation scenarios are specified in Supplementary Appendix B. We use nested loop plots to present the performance of all methods to illustrate how the simulation factors and parameter values influenced the estimates across all scenarios.[19, 20]

RESULTS

The key results of the simulation study are provided in the following subsections, structured by type of non-adherence. The results reported in this article are focused on methods performance in terms of unbiasedness including percentage bias (as a percentage of the truth) and absolute bias. The true values of the estimates across the 90 scenarios along with detailed performance results are reported in Supplementary Appendix D. Further plots showing the performance of methods in terms of the other measures (accuracy, coverage, empirical standard error, model-based standard error) are provided in Supplementary Appendix E.

Performance of Methods Across Implementation Non-Adherence Scenarios

The performance of methods in terms of percent bias across the implementation non-adherence scenarios is shown in Figure 2. G-methods (SNFTM, IPCW) performed the best across most scenarios although in scenarios with large treatment effect size (Scenarios 7-14 and 27-34) there is a modest increase in its percent bias. The level of non-adherence has some influence on percent bias, but g-methods generally handled the variation in non-adherence marginally better than other methods. PP performed best in 8 out of 38 scenarios and generally produced results close to g-methods in terms of bias. The levels of bias between IPCW, SNFTM and PP were often very similar.

ITT was always the worst-performing method as it produced a higher bias percent (as an overestimate) reaching more than 50% in some cases (Figure 2). ITT bias was amplified by a larger treatment effect size. Generally, levels of percent bias ranged between 7-37% for the IPCW (20% on average), 8-38% for SNFTM (20% on average), 8-38% for PP (20% on average) for PP, but between 20-75% for ITT (40% on average). Looking across all factors, it is clear that treatment effect size is one of the factors contributing to higher bias percentages for ITT, with large effect size increasing bias. Other factors contributing to bias include the type of survival time data-generating model.

In terms of absolute bias (Figure 3), g-methods produced a small bias of 0.019 years (7 days) on average (12 months follow-up) across all 38 implementation non-adherence scenarios, followed by PP that resulted in an average bias of 0.020 (7.4 days). In contrast, ITT resulted in a higher bias of 0.036 years (13.1 days) across the same implementation non-adherence scenarios.

Performance of Methods Across Persistence Non-Adherence Scenarios

Figure 4 illustrates methods performance in terms of percent bias across the persistence non-adherence scenarios. G-methods produced the least bias across 13 out of the 18 scenarios with PP producing better results in 5 scenarios with a small sample size. In all scenarios, the difference between g-methods and PP method is very small.

Small sample size has a negative impact on the performance of g-methods leading to higher bias compared to PP. G-methods and PP produced lower bias in direction of overestimation with bias percent around 20% although this is slightly higher (around 30%) in scenarios with a small sample size. ITT generated substantially higher bias, in most scenarios reaching up to 60% of the true value. Generally, levels of percent bias ranged between 9-36% for the IPCW (26% on average), 14-39% for SNFTM (26% on average), 14-36% for SNFTM (26% on average), but between 16-72% for ITT (47% on average).

In terms of absolute bias (Figure 5), g-methods produced a small bias of 0.023 years (8.3 days) on average (12 months follow-up) across all 18 persistence non-adherence scenarios. PP produced results that were very close to the g-methods with an average absolute bias of 0.023 (8.4 days). In contrast to a higher bias of 0.04 years (14.7 days) produced by ITT across the same scenarios.

Performance of Methods Across Initiation Non-Adherence Scenarios

Figure 6 illustrates percent bias across the 34 initiation non-adherence scenarios. SNFTM with g-estimation produced the smallest bias compared to the alternative methods in most scenarios. In contrast to the performance results in implementation and persistence non-adherence, PP did better than IPCW. Similar to the other types of non-adherence, the difference between g-methods and PP is very small. Bias performance fluctuated, largely dependent on the data-generating model, sample size and treatment effect size.

In contrast, ITT performance was worse than g-methods and PP in terms of percent bias across all but three scenarios. The direction of bias produced by all methods was in the positive region suggesting an overestimation, although the direction of bias changed to negative (suggesting an under-estimation) in some scenarios (Scenarios 83-86). This change was influenced by a combination of small sample size, large treatment effect size and time-dependent treatment effect. In scenarios with a large sample size (Scenarios 57-74), the average bias percentage produced by SNFTM was 14.7% in contrast with an average bias of 52.1% generated by the ITT analysis. PP resulted in a low bias percentage (close to g-methods) in the positive region of the nested loop plot indicating over-estimation in most scenarios.

SNFTM with g-estimation had better performance in scenarios with a large sample size with bias percent closer to zero in most cases (see Figure 6). In addition to the small sample size, the other main influencer of bias seems to be the treatment effect size with a larger effect size leading to higher bias, although g-methods and PP performed better than ITT in all scenarios. Generally, levels of percent bias ranged between -29 to 110% for the IPCW, -34 to 108% for SNFTM, -32 to 102% for PP, but between -18 to 200% for ITT. The negative percent bias was particularly seen in scenarios with small sample size and existence of time-dependent treatment effect (i.e., Scenarios 83-90).

Figure 7 presents performance as absolute bias with SNFTM producing the smallest bias of 0.035 years (12.8 days) on average compared to the alternative methods across the 34 initiation non-adherence scenarios with even smaller bias in scenarios with a large sample size (0.015 years [5.4 days]). This is followed by PP and IPCW that generated absolute bias of 0.036 (13.1) and 0.036 (13.2), respectively. In contrast, ITT was the worst-performing method as it resulted in a higher bias of up to 0.064 years (23.3 days) across the same scenarios.

DISCUSSION

The overall pattern of findings from this simulation study suggests that when there is non-adherence and we want to estimate what the treatment effect would have been had there been no non-adherence, estimates based on the ITT analysis will result in substantial bias. This is to be expected, because the ITT analysis does not address the estimand that we are interested in. The adjustment methods (IPCW, SNFTM, PP) all improve on this by reducing bias by about a half or more, even though with these better adjustment methods considerable bias may remain in some scenarios. This is likely to be linked to the presence of adequate data to be able to model the non-adherence process accurately.

The findings demonstrate the importance of adjusting the treatment effect for non-adherence in the context of time-to-event outcomes and HTA. Of note, non-adherence to medications is a major problem in healthcare, with significant consequences for both clinical outcomes and healthcare costs. HTAs generally compare incremental costs (IC) and incremental effectiveness (IE) of new technologies. In cases where the incremental costs are primarily driven by the cost of a drug, both the IC and the IE may be impacted by non-adherence. The focus of this simulation study was on the impact of non-adherence on treatment effectiveness, but in an HTA, costs would also need to be adjusted to reflect adherence levels expected in clinical practice.

In our simulation study we investigated scenarios where there was non-adherence in an RCT, and we wished to estimate what the treatment effect would have been had there been zero non-adherence. In reality, achieving zero non-adherence is not possible in most clinical settings and instead we would expect adherence to be lower in standard clinical practice than in RCTs, and therefore would wish to estimate the treatment effect under a lower level of adherence. Our study was designed to investigate whether the adjustment methods can accurately adjust for non-adherence to estimate the treatment effect under a different adherence level. For simplicity, the level we chose to adjust to was perfect adherence, but in reality, we would adjust to the level of adherence expected in standard clinical practice. The adjustment of effectiveness to other non-perfect adherence levels has been undertaken in a case study on kidney transplantation reported elsewhere. [25]

Another important research finding relates to PP analysis, which in many scenarios had a similar level of performance to that of the g-methods. This finding was unexpected as PP represents a very simple method of adjustment, and it seems related to the impact of the time-dependent confounder on graft survival times. In the simulated scenarios, patients with worse prognostic characteristics (i.e., high BMI and aged below 24 years) were more likely to non-adhere to the dosing regimen, but these patients were also more likely to experience the graft loss event. Therefore, over time, fewer of these patients remain in the dataset (and therefore cannot non-adhere) – at these later time-points better prognosis patients remain in the dataset and therefore over time more good prognosis patients have the possibility of non-adhering. For this reason, over the time-course of the simulated datasets, a mixture of poor prognosis and good prognosis patients non-adhere, and the relationship between non-adherence and prognostic characteristics is weakened over time. This has two important implications – firstly, it may be difficult for the g-methods to accurately model the non-adherence process (especially when only three time-points were simulated for each dataset) meaning that these methods produce appreciable bias; secondly, because there is not a strong relationship between prognosis and adherence, the PP approach is not very biased. In order to examine this issue further, more complex DGMs may be required, and this could be an important area for future research. The DGMs should include scenarios with patient characteristics that are more realistic (e.g., calibrated with existing RCT data) and potentially additional prognostic factors could be included. However, our simulations were sufficiently complex to assess the performance of g-methods and PP compared to ITT analysis across all types of non-adherence. The study provides clear evidence in favour of g-methods against ITT analyses.

Existing evidence from relevant simulation studies showed comparable results, although the design of those simulation studies was less comprehensive than the work described here. Cain and Cole, 2009[21] published the first simulation evidence that compared IPCW and ITT methods in adjusting for non-adherence in the context of survival analysis. The study assessed methods performance in three scenarios defined by different levels of non-adherence.[21] Bias and MSE were reported as performance measures with findings showing that IPCW was the best-performing method in terms of unbiasedness. The paper reported that bias and imprecision increased as the level of non-adherence increased. This is similar to findings from our current simulation study, although the magnitude of bias differs, potentially due to the different estimands, DGMs, non-adherence metrics and other parameter values.

Other existing simulation evidence includes a study reported by Zhang et al., 2011[22] in which they compared the same methods (IPCW versus ITT) for adjusting for persistence non-adherence. The simulated datasets included baseline covariates, time-dependent confounders, censoring of event time, and time-varying non-adherence. The key findings are similar to those reported by Cain and Cole, with IPCW generating the best performance in terms of unbiasedness and coverage. Although a different study design was used (including the estimand), the findings of this study are similar to our simulation study, in which IPCW outperforms the ITT method. However, the magnitude of coverage differs and this is likely to be due to differences in DGMs and sample size. Our study findings show that a smaller sample size leads to poor performance, which the other studies were unable to reveal by their use of arbitrary sample sizes that failed to capture realistic trial designs. It was not possible to compare the findings related to the PP method because the two relevant existing studies did not include PP analysis in the alternative methods assessed.

The g-methods evaluated here have also been tested in several simulation studies for adjusting effectiveness in the presence of treatment switching.[17, 23, 24] The context is different, but the methods considered and the concepts are similar, and these have generally shown that IPCW is preferable to PP, but both can be very biased when sample sizes are small or if the population being allocated weights is very small. The findings from these simulation studies show that g-methods were superior to simple methods in terms of performance using a range of performance measures including bias and coverage. For instance, Latimer et al.[17] published findings from a simulation study aimed to assess the performance of IPCW, RPSFTM and ITT (among other methods). The authors simulated RCT datasets in the presence of treatment switching with a time-to-vent outcome and time-dependent confounding. The primary estimand was RMST in the control group that would have been observed in the absence of treatment switching. The findings from the simulations demonstrated that both g-methods (IPCW and RPSFTM) performed well in terms of handling the effects of time-dependent confounding. However, IPCW struggled in situations with high percentages of switching and a modest sample size.

Fiorentino et al. [26] recently published findings from a systematic review of methods to handle adherence in RCTs and found that only 13% of trials account for adherence, highlighting that no causal inference methods were used. The paper concluded that increased awareness to adjust for adherence in the analyses of RCT data using statistical methods is needed. Our current paper contributes to addressing this need by providing new simulation evidence on the performance of non-adherence adjustment methods across a range of 90 scenarios.

Strengths and Limitations

The simulation study design was based on recent international guidelines for planning simulation studies and followed a pre-specified study protocol.[7, 10, 11] The study included the simple censoring PP method, which was not considered in the existing comparable studies, and this has produced interesting findings as discussed above. Moreover, robust DGMs were used to simulate biologically plausible survival data with baseline covariates, a time-dependent confounder, time-varying non-adherence, and a robust randomisation assignment procedure. The relationships between prognostic variables, non-adherence and the survival outcome followed a DAG. The user-written baseline hazard functions used for simulating survival data in a delayed entry model allowed for simulating a range of RCT datasets for testing the alternative non-adherence adjustment methods.

There are also several limitations associated with this simulation study. The potentially higher number of baseline and time-dependent confounders in real data and interactions between multiple confounders, non-adherence, and survival outcome might lead to different outcomes to those generated in simulated datasets and this might influence performance. Non-convergence problems associated with the DGMs should also be considered as a limitation. However, enough successful simulations were achieved to assess the alternative non-adherence adjustment method for each scenario assessed. Across all our simulated scenarios, all methods converged in more than 92% of simulations.

As another limitation, the assumption of "no unmeasured confounding" was not assessed in this simulation study. This is because only one time-dependent confounder was simulated, thereby making it impossible to run analyses with fewer covariates. PP analysis could be considered as a baseline for this. The method does not correct for differences between non-adherers and adherers, and as such is equivalent to IPCW without the inclusion of any covariates.

The RPSFTM was identified as appropriate for adjusting for real-world adherence levels, but its performance was not assessed in our simulation study. The RPSFTM only works to adjust for non-adherence in single-arm studies or studies with placebo-control arms; however, our simulation study was designed to assess the performance of adjustment methods in RCTs with two active treatments, and hence, the method was excluded. Similarly, the PKPD method was identified as appropriate; however, it was excluded from the simulations because it requires a different study design in terms of DGMs in order to directly compare it with the g-methods. Further research is recommended to assess the performance of RPSFTM and PKPD methods in a direct comparison with IPCW and SNFTM using a well-conducted simulation study.

Conclusions

The simulation study demonstrates that g-methods and PP analysis provide more accurate estimates of the treatment effect than the ITT analysis, when adjustment for non-adherence is required. However, considerable bias may remain in some scenarios. When real-world adherence is expected to differ from adherence observed in a trial, adjustment methods should be used to provide estimates of real-world effectiveness.

Acknowledgements

"See Title Page for Acknowledgements"

Conflict of Interests

“See Title Page for other declaration of conflicting interests”

REFERENCES

1. Hughes D, Charles J, Dawoud D, et al. Conducting Economic Evaluations Alongside Randomised Trials: Current Methodological Issues and Novel Approaches. *Pharmacoeconomics* 2016. DOI: 10.1007/s40273-015-0371-y.

2. Cleemput I, Kesteloot K and DeGeest S. A review of the literature on the economics of noncompliance. Room for methodological improvement. *Health Policy* 2002; 59: 65–94. DOI: S0168-8510(01)00178-6.

3. Barber N. Should we consider non-compliance a medical error? *Qual Saf Health Care* 2002; 11: 81–84.

4. Vermeire E, Hearnshaw H, Royen PV and Denekens J. Patient adherence to treatment: three decades of research. A comprehensive review. *J Clin Pharm Ther* 2001; 26: 331–342.

5. Hughes DA, Bagust A, Haycox A and Walley T. Accounting for noncompliance in pharmacoeconomic evaluations. *Pharmacoeconomics* 2001; 19: 1185-1197. DOI: 10.2165/00019053-200119120-00001.

6. Alshreef A, Latimer N, Tappenden P, et al. Statistical Methods for Adjusting Estimates of Treatment Effectiveness for Patient Nonadherence in the Context of Time-to-Event Outcomes and Health Technology Assessment: A Systematic Review of Methodological Papers. *Med Decis Making* 2019; 39: 910–925. DOI: 10.1177/0272989X19881654.

7. Vrijens B, Geest SD, Hughes DA, et al. A new taxonomy for describing and defining adherence to medications. *Br J Clin Pharmacol* 2012; 73: 691–705.

8. Andrzejczyk A, Clyne W, Geest SD, et al. *Ascertaining barriers for compliance: policies for safe, effective and cost-effective use of medicines in Europe*. 2012.

9. Muszbek N, Brixner D, Benedict A, et al. The economic consequences of noncompliance in cardiovascular disease and related conditions: a literature review. *Int J Clin Pract* 2008; 62: 338–351. DOI: 10.1111/j.1742-1241.2007.01683.x.

10. Burton A, Altman DG, Royston P and Holder RL. The design of simulation studies in medical statistics. *Stat Med* 2006; 25: 4279–4292. DOI: 10.1002/sim.2673.

11. Morris TP, White I and Crowther MJ. Using simulation studies to evaluate statistical methods. *Stat Med* 2019; 38: 2074-2102. DOI: 10.1002/sim.8086.

12. Pinsky BW, Takemoto SK, Lentine KL, et al. Transplant Outcomes and Economic Costs Associated with Patient Noncompliance to Immunosuppression. *American Journal of Transplantation* 2009; 9: 2597-2606. DOI: 10.1111/j.1600-6143.2009.02798.x.

13. Kennedy C and Mann CB. *"RANDOMIZE: Stata module to create random assignments for experimental trials, including blocking, balance checking, and automated rerandomization*. 2015. Boston College Department of Economics.

14. Crowther MJ and Lambert PC. Simulating biologically plausible complex survival data. *Stat Med* 2013; 32: 4118–4134. DOI: 10.1002/sim.5823.

15. Ekberg H, Tedesco-Silva H, Demirbas A, et al. Reduced Exposure to Calcineurin Inhibitors in Renal Transplantation. *N Engl J Med* 2007; 357: 2562-2575. DOI: 10.1056/NEJMoa067411.

16. Royston P and Parmar MK. Restricted mean survival time: an alternative to the hazard ratio for the design and analysis of randomized trials with a time-to-event outcome. *BMC Med Res Methodol* 2013; 13: 1-15.

17. Latimer N, White I, Tilling K and Siebert U. Improved two-stage estimation to adjust for treatment switching in randomised trials: g-estimation to address time-dependent confounding. *Stat Methods Med Res* 2020; 29: 2900–2918. DOI: 10.1177/0962280220912524.

18. White IR. simsum: Analyses of simulation studies including Monte Carlo error. *Stata J* 2010; 10: 369-385. DOI: 10.1177/1536867X1001000305.

19. Latimer NR, Abrams KR and Sieber U. Two-stage estimation to adjust for treatment switching in randomised trials: a simulation study investigating the use of inverse probability weighting instead of re-censoring. *BMC Med Res Methodol* 2019; 19: 1-19. DOI: 10.1186/s12874-019-0709-9.

20. Rücker G and Schwarzer G. Presenting simulation results in a nested loop plot. *BMC Med Res Methodol* 2014; 14: 1-8.

21. Cain LE and Cole SR. Inverse probability-of-censoring weights for the correction of time-varying noncompliance in the effect of randomized highly active antiretroviral therapy on incident AIDS or death. *Stat Med* 2009; 28: 1725-1738. DOI: 10.1002/sim.3585.

22. Zhang M, Tsiatis AA, Davidian M, et al. Inference on treatment effects from a randomized clinical trial in the presence of premature treatment discontinuation: the SYNERGY trial. *Biostatistics* 2011; 12: 258-269. DOI: 10.1093/biostatistics/kxq054.

23. Latimer NR, Abrams KR, Lambert PC, et al. Assessing methods for dealing with treatment switching in clinical trials: A follow-up simulation study. *Stat Methods Med Res* 2018; 27: 765–784. DOI: 10.1177/0962280216642264.

24. Latimer NR, Abrams K, Lambert P, et al. Adjusting for treatment switching in randomised controlled trials – A simulation study and a simplified two-stage method. *Stat Methods Med Res* 2017; 26: 724–751. DOI: 0.1177/0962280214557578.

25. Alshreef A, Latimer N, Tappenden P, et al. A Case Study Assessing the Cost-Effectiveness of Maintenance Immunosuppressive Therapy for Kidney Transplantation in Adults: Accounting for Non-Adherence Using Data from a Randomised Trial and the Real World within a Decision-Analytic Model. *Value Health* 2022; 25: S529. DOI: 10.1016/j.jval.2022.04.1266.

26. Fiorentinoa F, Nohpal de la Rosab C, Day E., et al. Statistical methods for handling compliance in randomized controlled trials of device interventions: a systematic review. *Journal of Clin. Epi* 2022; 152: 226-237. DOI: 10.1016/j.jclinepi.2022.09.015.

TABLES

Table 1: Factors Included in The Simulation Study Scenarios

Factor level	Sample size	Type of non-adherence	Non-adherence	Graft survival time data-generating model	Relationship between treatment effect and adherence	Time-dependent treatment effect	Treatment effect size
1	Large	Implementation	High	Standard parametric survival model (PSM), Weibull distribution	Strong	Yes	Large
2	Small	Persistence	Low	Two-component PSM, Weibull-Weibull (Mixture) distribution	Weak	No	Small
3	-	Initiation	-	-	-	-	-

BOXES

Box 1: Simulation Process and Outputs

1. RCT datasets were simulated using 1 million iterations with prognostic baseline and time-dependent confounding and graft survival outcomes. The average treatment effect (difference in RMSTs) generated from the analysis of these datasets represents the "true" treatment effect.

2. RCT datasets were simulated using 1900 iterations and a similar DGM specified in (1) with the only difference being the presence of non-adherence. In these datasets, the probability of non-adherence was associated with the prognostic baseline (age) and time-dependent (BMI) characteristics. This step was repeated for 90 scenarios with different types and levels on non-adherence.

3. The alternative four methods were applied to adjust for non-adherence in the dataset simulated in (2) to estimate the adherence-adjusted treatment effect. The key estimates generated from this step include the difference in RMSTs, standard errors and confidence intervals for the difference in RMSTs, and indicators for model convergence.

4. Each estimates dataset was analysed using the “simsum” Stata command for assessing the performance of each method which was computed using results from the performance measures as a percentage of the truth. In addition, the Monte Carlo standard error was produced to assess simulation uncertainty in each scenario.

FIGURES

Figure 1

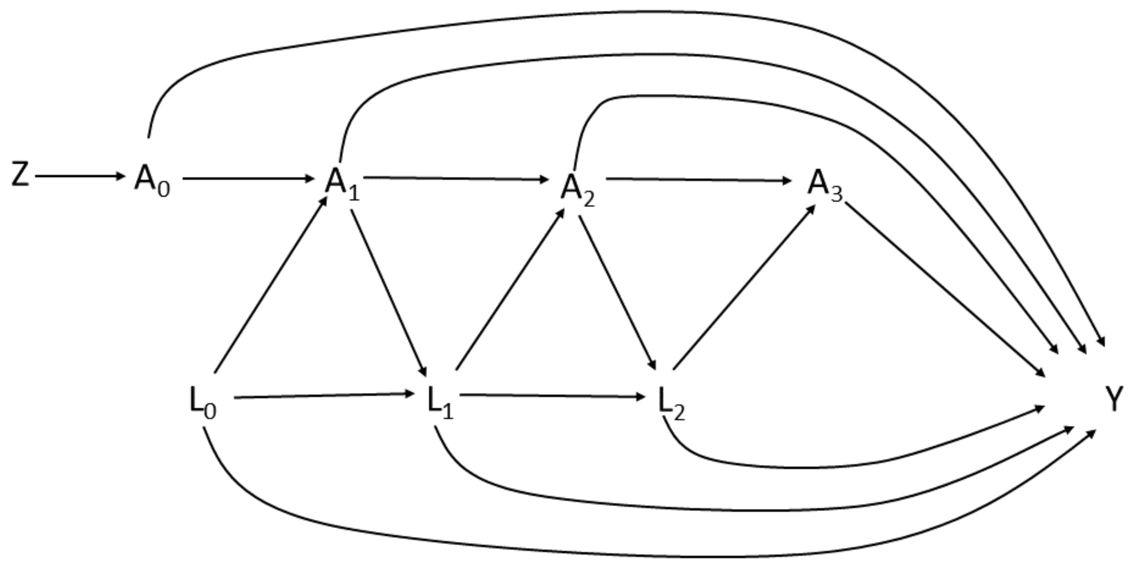


Figure 2

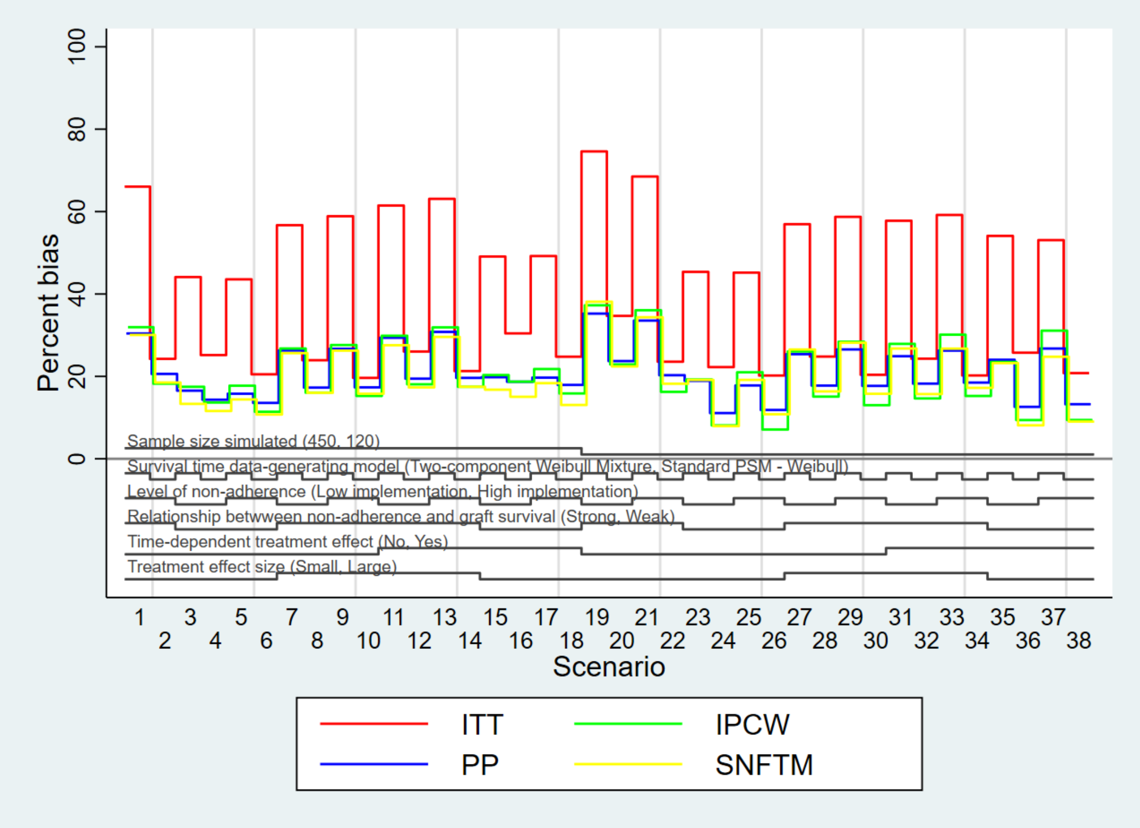


Figure 3

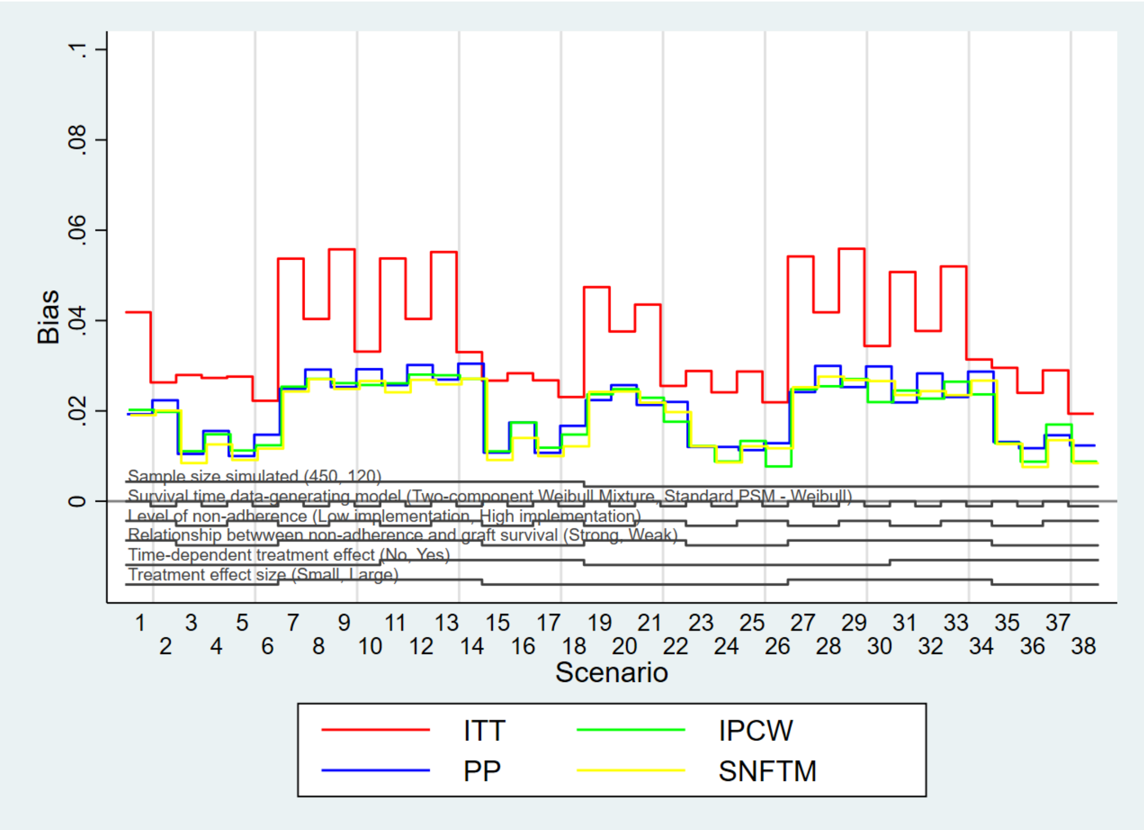


Figure 4

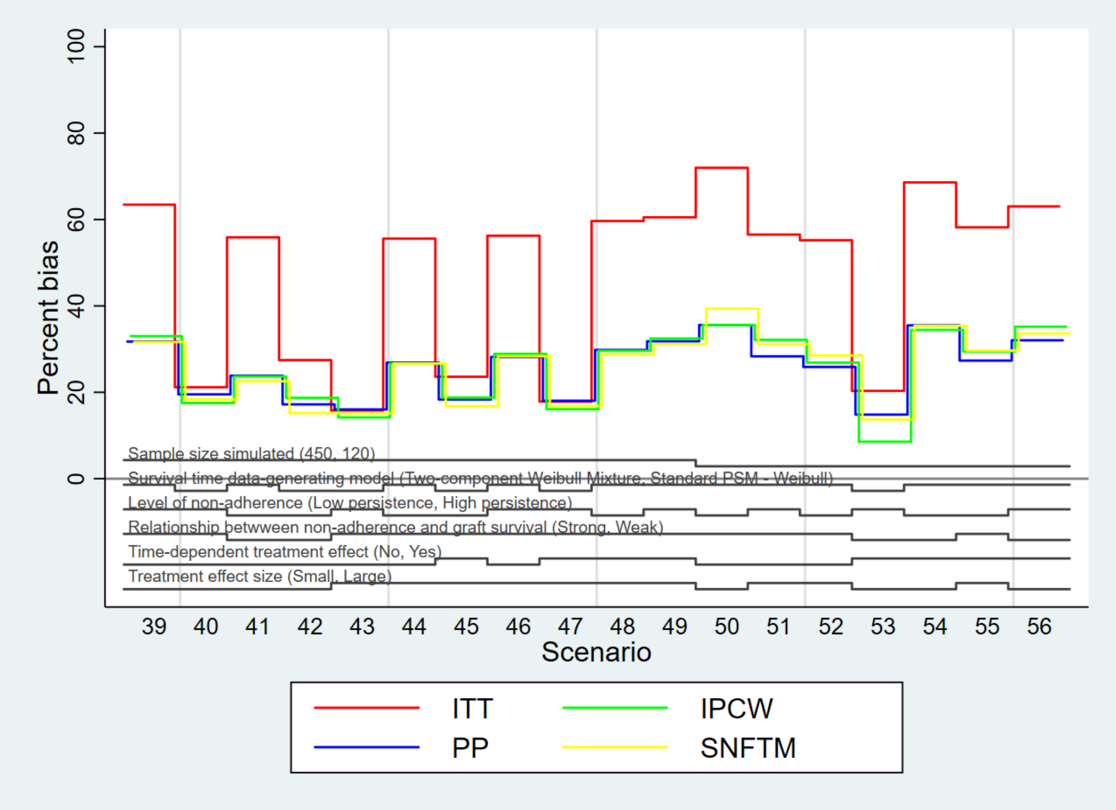


Figure 5

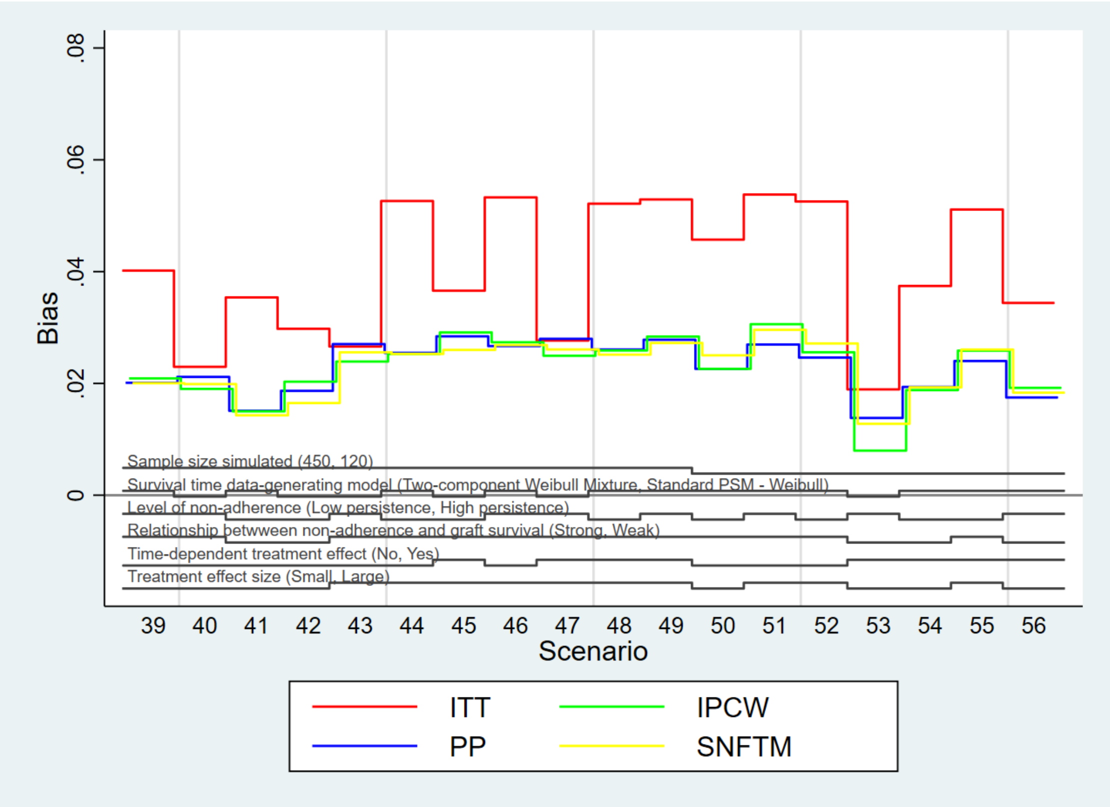


Figure 6

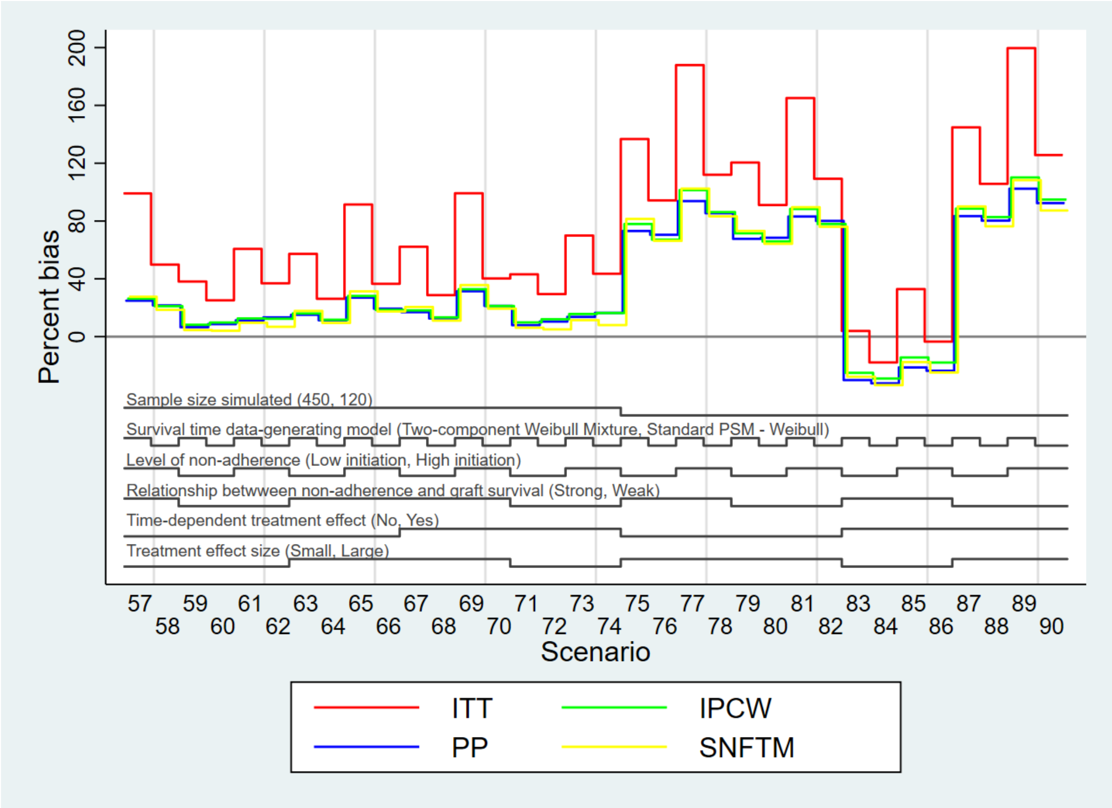


Figure 7

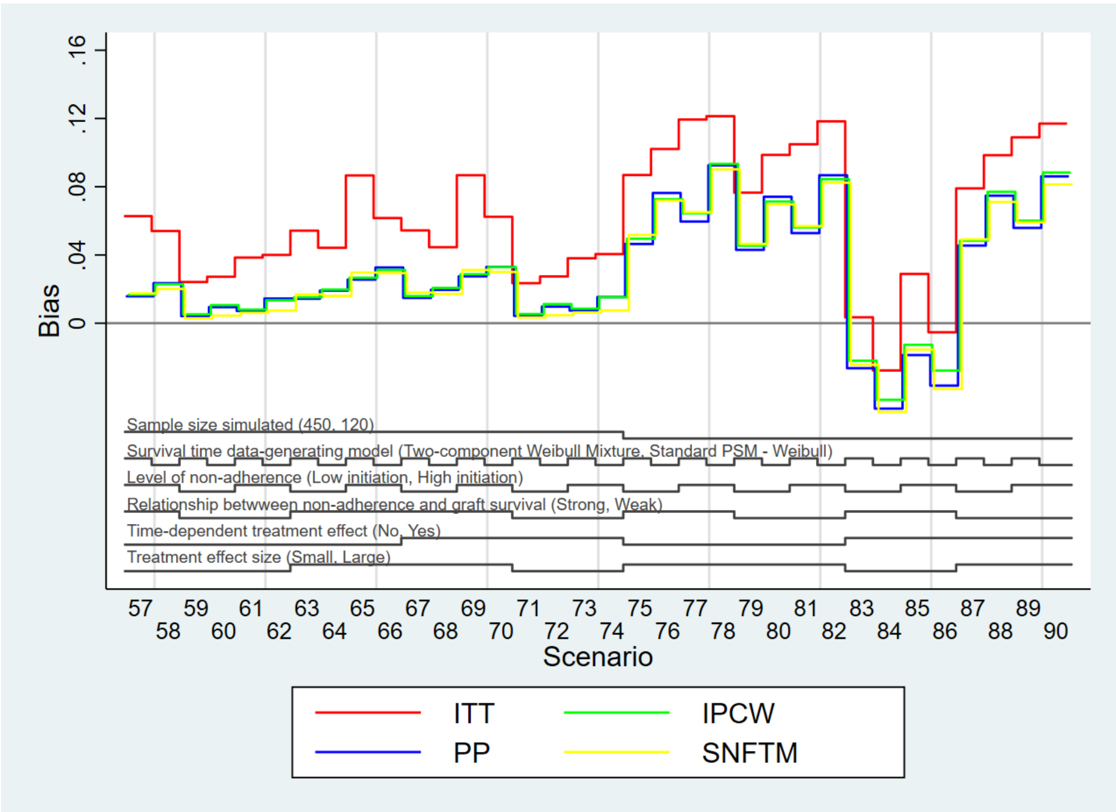


FIGURE LEGENDS

Figure 1: A directed-acyclic graph (DAG) representing variable relationships in the data-generating model for implementation non-adherence. Z, is randomisation; L_0, is a vector of baseline covariates that includes age and values of time-varying covariates measured at baseline; L_1, L_2, updated time-varying covariate (BMI) at 4 and 8 months, respectively; A_0, A_1, A_2, A_3, time-varying non-adherence at baseline, between baseline and 4 months, 4- 8 months and 8-12 months, respectively.

Figure 2: Percentage bias in the estimation of the difference in RMSTs across implementation non-adherence scenarios. PSM, parametric survival model; ITT, intention-to-treat; PP, per-protocol; IPCW, inverse probability of censoring weighting; SNFTM, structural nested failure time model.

Figure 3: Bias in the estimation of the difference in RMSTs across implementation non-adherence scenarios**.** PSM, parametric survival model; ITT, intention-to-treat; PP, per-protocol; IPCW, inverse probability of censoring weighting; SNFTM, structural nested failure time model.

Figure 4: Percentage bias in the estimation of the difference in RMSTs across persistence non-adherence scenarios**.** PSM, parametric survival model; ITT, intention-to-treat; PP, per-protocol; IPCW, inverse probability of censoring weighting; SNFTM, structural nested failure time model.

Figure 5: Bias in the estimation of the difference in RMSTs across persistence non-adherence scenarios**.** PSM, parametric survival model; ITT, intention-to-treat; PP, per-protocol; IPCW, inverse probability of censoring weighting; SNFTM, structural nested failure time model

Figure 6: Percentage bias in the estimation of the difference in RMSTs across initiation non-adherence scenarios. PSM, parametric survival model; ITT, intention-to-treat; PP, per-protocol; IPCW, inverse probability of censoring weighting; SNFTM, structural nested failure time model.

Figure 7: Bias in the estimation of the difference in RMSTs across initiation non-adherence scenarios**.** PSM, parametric survival model; ITT, intention-to-treat; PP, per-protocol; IPCW, inverse probability of censoring weighting; SNFTM, structural nested failure time model.

APPENDICES

Supplementary Appendix A: Detailed Data-Generating Mechanisms

1. Parameter Values and Distributions

To generate biologically plausible RCT datasets, the parameters' values for the data-generating mechanisms (DGMs) were specified. The following factors were varied to specify the scenarios simulated: sample size, non-adherence metrics, baseline hazard function for the survival-time data-generating model, the relationship between adherence level and graft survival, time-dependent treatment effects and treatment effect size. The parameter values for each factor that were used to simulate the dataset are specified in the following sections (Sections 1.1 – 1.6).

1.1 Sample Size

To specify the number of observations in the simulated datasets, the sample size for large studies (n=450) was assumed based on the 75th percentile sample size among 40 trials (Table 1). These were clinical trials conducted in the area of maintenance immunosuppression after kidney transplantation, as identified by a published systematic review and included in a network meta-analysis (NMA).[1]

A small sample size (n=120) was assumed for some scenarios which were based on the 25th percentile sample size among the same list of trials (Table 1). Initially, I planned to use a sample size of 90 but based on testing, that resulted in a very high number of failed simulations. This was largely due to non-convergence associated with the nature of how some non-adherence adjustment methods work (i.e., g-methods). This issue is discussed later in this thesis.

Table 2: Sample size for clinical trials assessed maintenance immunosuppression after kidney transplantation

Author (Year)	n
Grinyo 2009	1529
Vincenti 2010	686
Waller 2002	102
Laskow 1996	120
Baboolal 2002	51
Campos 2002	166
Margreiter 2002	560
Sadek 2002	477
Mayer 1997	448
Tricontinental MMF renal study 1996	497

Author (Year)	n
Yang 1999	60
Schaefer 2006	80
Hardinger 2005	200
Rowshani 2006	126
Weimer 2006	81
Tedesco-Silva 2010	783
Barsoum 2007	113
Anil Kumar 2005	150
Mendez 2005	361
Sampaio 2008	100
Martinez-Mier 2006	41
Nafar 2012	100
Anil Kumar 2008	200
Vincenti 2005	218
Ferguson 2011	89
Flechner 2002	61
Vacher-Coponat 2012	289
Büchler 2007	145
Charpentier 2003	83
Merville 2004	71
Durrbach 2010	578
Lorber 2005	583
Bertoni 2011	106
Gallon 2006	83
Guba 2010	140
Lebranchu 2009	192
Vítko 2005	588
Larson 2006	162
Chadban 2013	126
Flechner 2011	450

1.2 Non-Adherence Metrics

Non-adherence was simulated using binary covariates for implementation, persistence or initiation (depending on the scenario specifications). Initiation and persistence use a binary variable that reflects the non-adherence event at a particular time point; whilst implementation reflects non-adherence to the prescribed regimen during a particular time interval (e.g., 4 to 8 months).

For implementation, non-adherence is often measured by the coefficient of variation (CV%), which is a validated measure of adherence in kidney transplant recipients that has been demonstrated to be associated with graft loss.[2, 3] A higher CV% indicates a more erratic level of medication-taking behaviour, and therefore, a higher level of non-adherence. At least three data points of drug concentration levels are needed to calculate the CV% for each individual patient for each time interval, which then needs to be combined with a cut-off point for CV% above which the patient will be categorised as non-adherent. Implementation non-adherence was simulated as a binary variable without simulating concentration levels and/or CV% and then converting it to a binary variable. Implementation non-adherence is expected to be more prevalent, but to have less impact on the graft survival outcome. Therefore, that is what I simulated in the implementation scenarios, by changing the numbers used for non-adherence probability and the effect of non-adherence on graft survival times in the DGMs.

The probabilities of non-adherence at each time interval (% of non-adherence) were simulated such that the overall non-adherence patterns are classified as high/low. These were numerically defined as relative values depending on the type of non-adherence (e.g., 10% low implementation non-adherence and 40% high implementation non-adherence). The probability of non-adherence in the control arm was different to the experimental arm. This mimics the usual pattern of non-adherence seen in clinical trials. The values for the probability of non-adherence were assumed with values varied depending on follow-up time points and the type of non-adherence evaluated in each set of scenarios (See Appendix C for an example of these parameter values).

1.3 Baseline Hazard Function

Two DGMs (Standard parametric survival model (PSM) with Weibull distribution and Two-component parametric survival model with Weibull-Weibull (Mixture) distribution) were used to generate graft survival times (Figures 1-2). For each model, a user-defined log hazard function with polynomial fractions and delayed entry was specified.[4] The shape of the KM survival curves generated by the standard PSM was sharply decreasing in hazards as shown from the analysis of a simulated dataset with 1000 observations. The shape of the KM survival curves generated by the mixture Weibull-Weibull model with similar patient characteristics is shown in Figure 13. Despite the similar patient characteristics, the mixture model produces survivor functions that drop a lot more slowly, so may be considered to represent less severe disease. The mixture model mimics graft survival curves observed in clinical trials conducted in kidney transplantation. The two models were used in this simulation to boost the transferability of findings beyond kidney transplantation as shape parameters for survival data varies across disease areas.

Figure 8: KM curves using standard parametric survival model with Weibull distribution

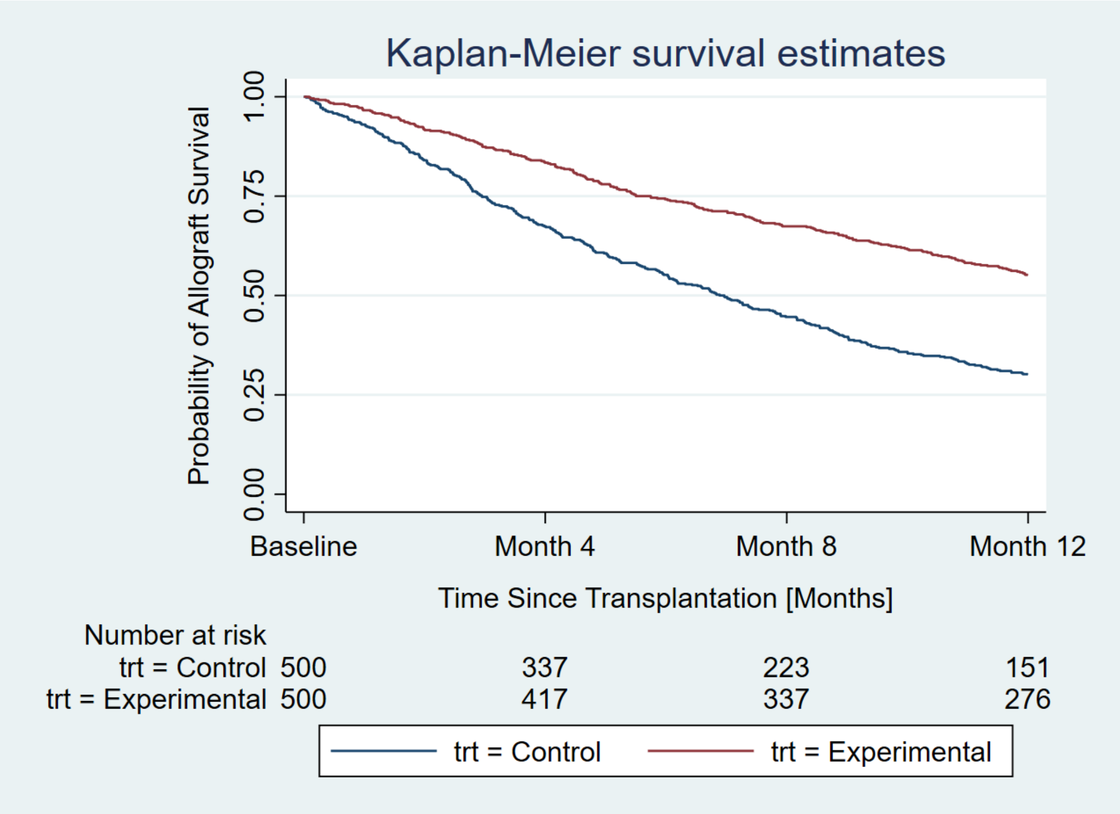
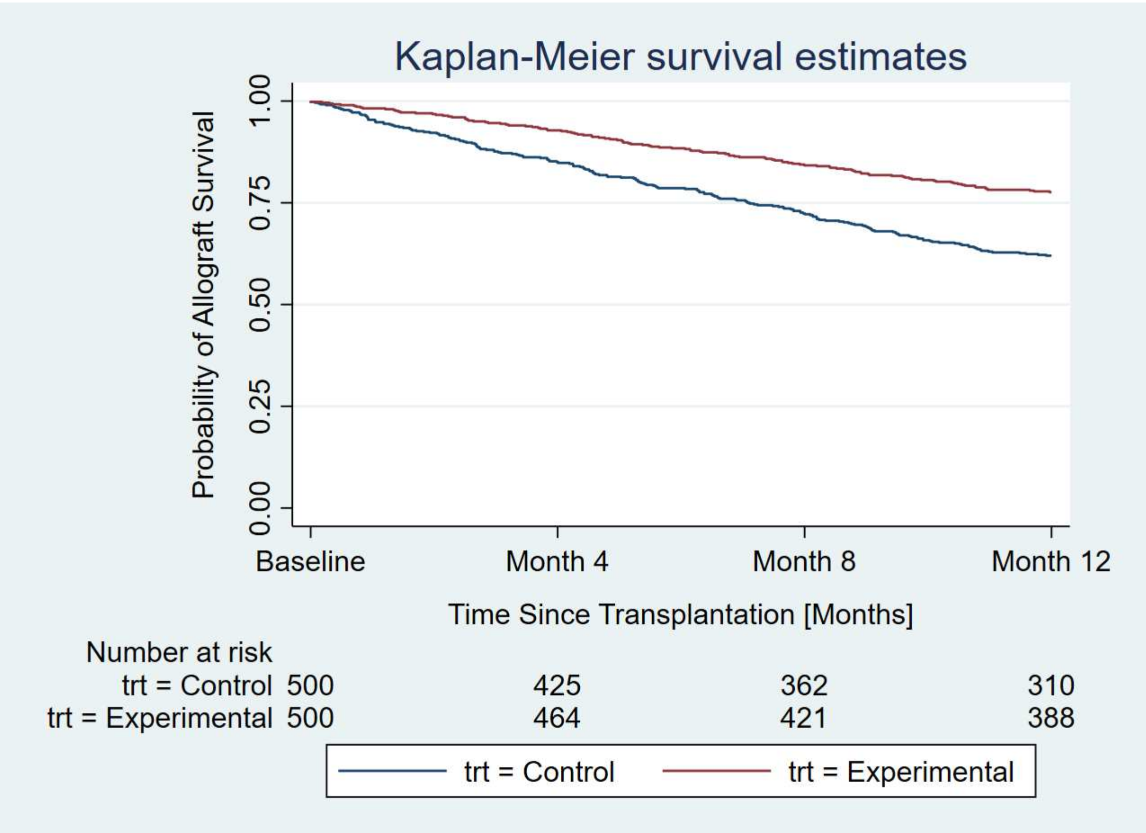


Figure 9: KM curves using two-component parametric survival model with Weibull-Weibull (Mixture) distribution



For the standard PSM, the parameter values used in the user-defined log hazard function were: *loghazard(-1.2:*0.2:*#t:^(0.2:-1)*. For the two-component mixture model, the parameters values for the hazard function were: *loghazard(-1.2 :+ 0.2:*#t :- 0.03:*#t:^0.5 :+ 0.05:*#t:^-0.5).* To implement these complex hazard functions, the Mata code was combined with the *survsim* command in Stata in a way that allowed for incorporating time-dependent covariates and non-adherence. The *"moremata"* Stata package was used to apply the models. The technical details and the full code used to simulate the datasets are provided in Appendix (G).

1.4 Relationship Between Non-Adherence and Graft Survival

A parameter was used to represent the correlation coefficient as a measure of how strong the relationship between non-adherence and the time-to-event outcome (time-to-graft loss) was. The value of this parameter was specified as a coefficient within the survsim model such that the relationship is classed as "strong" or "weak" depending on the scenario. For a strong relationship, a value of 0.40 was used across implementation non-adherence scenarios. For persistence and initiation non-adherence, higher values were used ranging between 0.42 to 0.55 to simulate a stronger impact on graft survival. For the weak relationship, a value of 0.22 was used in implementation non-adherence scenarios with alternative values ranging between 0.22 to 0.38 used in persistence and initiation non-adherence scenarios. These parameter values were assumed to achieve the desired Kaplan-Meier (KM) survival curves and hazard ratios (HRs). The strength values (strong/weak) are relative and these were determined using simple regression analysis on the simulated datasets. Different values of coefficients were used depending on the type of non-adherence to reflect different impacts on the survival outcome as discussed in Section 2.

1.5 Time-Dependent Treatment Effect

The time-dependent treatment effect was incorporated in some scenarios as specified in Appendix C. In these scenarios, the parameter value was 0.15 allowing for a 15% time-dependent linear reduction in the effect of the treatment. This value was assumed to achieve the desired KM curves and HRs. Therefore, scenarios with no time-dependent treatment effect assumed a 0% time-dependent reduction in treatment effect, and constant HR.

1.6 Treatment Effect Size

For generating the RCT datasets, the treatment effect size was specified based on HRs. Graft survival times were simulated such that the generated HR is around 0.55 indicating a beneficial treatment effect reducing the graft loss event rate by 45%. This large treatment effect is representative of comparing a very effective drug to a less effective drug (e.g., tacrolimus versus standard-dose cyclosporine regimens in some scenarios). In scenarios where a small/moderate treatment effect size was simulated, an HR of around 0.70 was generated representing other comparisons such as low-dose cyclosporine versus standard-dose cyclosporine.

2. Application of Coefficient Values Within Scenarios

To simulate the datasets, a range of other parameters were specified in the form of coefficient values incorporated into the simulation program. These include coefficients for generating baseline and time-dependant covariates, time-varying non-adherence and graft survival times. In the simulated datasets, both baseline and time-dependent confounders (age and BMI) were included as covariates. The coefficient values used within the simulation program for generating time-varying non-adherence, baseline covariates, time-dependent confounders and graft survival times were kept constant across simulations. The rationale is to focus on varying the values of the key factors (specified in Table 1 in the main body of the paper) to evaluate their influence on methods performance.

To explain how the simulation program is implemented in Stata Software, let us take Scenario 2 as an example. The parameter values and distributions specified for generating the datasets in this scenario are presented in Table 2.

Table 3: Parameter values for simulated RCT datasets - Scenario 2

Parameter	Value for Scenario 2	Distribution/function	Source of value
Sample size	450	-	Analysis based on data from Table 1 above
Age	18-24 (55%), 25-75 (45%)	Conditional random variable	Assumed
Treatment group	0= Control, 1= Experimental	Randomly assigned in 1:1 ratio	-
Non-adherence - implementation: 0-4 months	Control: 30% if (age ≤24 & hBMI0=1) 20% if (age >24 & hBMI0=1) 10% if (age ≤24 & hBMI0=0) 5% if (age >24 & hBMI0=0) Experimental: 22.5% if (age ≤24 & hBMI0=1) 15% if (age >24 & hBMI0=1) 10% if (age ≤24 & hBMI0=0) 5% if (age >24 & hBMI0=0)	Binomial random variable	-
Non-adherence - implementation: 4-8 months	Control: 60% if (age ≤24 & hBMI0=1) 40% if (age >24 & hBMI0=1) 20% if (age ≤24 & hBMI0=0) 10% if (age >24 & hBMI0=0) Experimental: 45% if (age ≤24 & hBMI0=1) 30% if (age >24 & hBMI0=1) 15% if (age ≤24 & hBMI0=0) 7.5% if (age >24 & hBMI0=0)	Binomial random variable	-

Parameter	Value for Scenario 2	Distribution/function	Source of value
Non-adherence - implementation: 8-12 months	Control: 60% if (age ≤24 & hBMI0=1) 40% if (age >24 & hBMI0=1) 20% if (age ≤24 & hBMI0=0) 10% if (age >24 & hBMI0=0) Experimental: 45% if (age ≤24 & hBMI0=1) 30% if (age >24 & hBMI0=1) 15% if (age ≤24 & hBMI0=0) 7.5% if (age >24 & hBMI0=0)	Binomial random variable	-
Probability of high BMI (≥30) at baseline	(1,0.60)	Binomial random variable	-
Probability of high BMI (≥30) at Month 4	Control & Experimental: 90% if (age ≤24 & hBMI0=1) 30% if (age >24 & hBMI0=1) 60% if (age ≤24 & hBMI0=0) 20% if (age >24 & hBMI0=0)	Binomial random variable	-
Probability of high BMI (≥30) at Month 8	Control & Experimental: 90% if (age ≤24 & hBMI0=1) 30% if (age >24 & hBMI0=1) 60% if (age ≤24 & hBMI0=0) 20% if (age >24 & hBMI0=0)	Binomial random variable	-
Baseline hazard function	loghazard(-1.2 :+ 0.2:*#t :- 0.03:*#t:^0.5 :+ 0.05:*#t:^-0.5)	User-written hazard function using a two-component parametric survival model - Weibull-Weibull (Mixture) distribution	
Coefficients for generating graft survival time	$\beta_0 L_0$ (Age) = 0.25 $\beta_0 L_0$ (hBMI0) = 0.35 ($\beta_1 L_1$(hBMI1) = 0.35 ($\beta_2 L_2$(hBMI2) = 0.35 (β_{1a1} (A1) = 0.40 (β_{2a2} (A2) = 0.40 (β_{3a3} (A3) = 0.40 Treatment effect= -0.75 Time-dependent effect= 0	Implemented within the *"survsim"* model in Stata	
Administrative censoring (End of study) in Years	1.0		

β_0, the coefficient for baseline covariates and the value of time-dependent covariates at baseline (L_0); β_1, the coefficient for time-dependent covariates at 4 months (L_1); β_2, the coefficient for time-dependent covariates at 8 months (L_2); β_{1a1}, β_{2a2} and β_{3a3}, the coefficients for implementation non-adherence between baseline at 4 months (A1), 4 to 8 months (A2) and 8 to 12 months (A3); hBMI0, high Body Mass Index at baseline; hBMI1, high Body Mass Index at Month 4; hBMI1, high Body Mass Index at Month 8. A1, Implementation non-adherence between baseline and 4 months; A2, Implementation non-adherence between 4 and 8 months; A3, Implementation non-adherence between 8 and 12 months.

The parameter values presented in Table 2 were incorporated into the simulation program and *survsim* model to produce graft survival times in the absence of non-adherence. The Kaplan-Meier survival curves produced from one simulated dataset with a sample size of 2000 and perfect adherence using the relevant parameter values (Table 2) are presented in Figure 3. This step was run 1 million times to produce the truth for one scenario in the full simulation program. Figure 4 shows KM survival curves with non-adherence incorporated.

Figure 10: Graft survival curves from the simulated dataset in the absence of non-adherence - Scenario 2

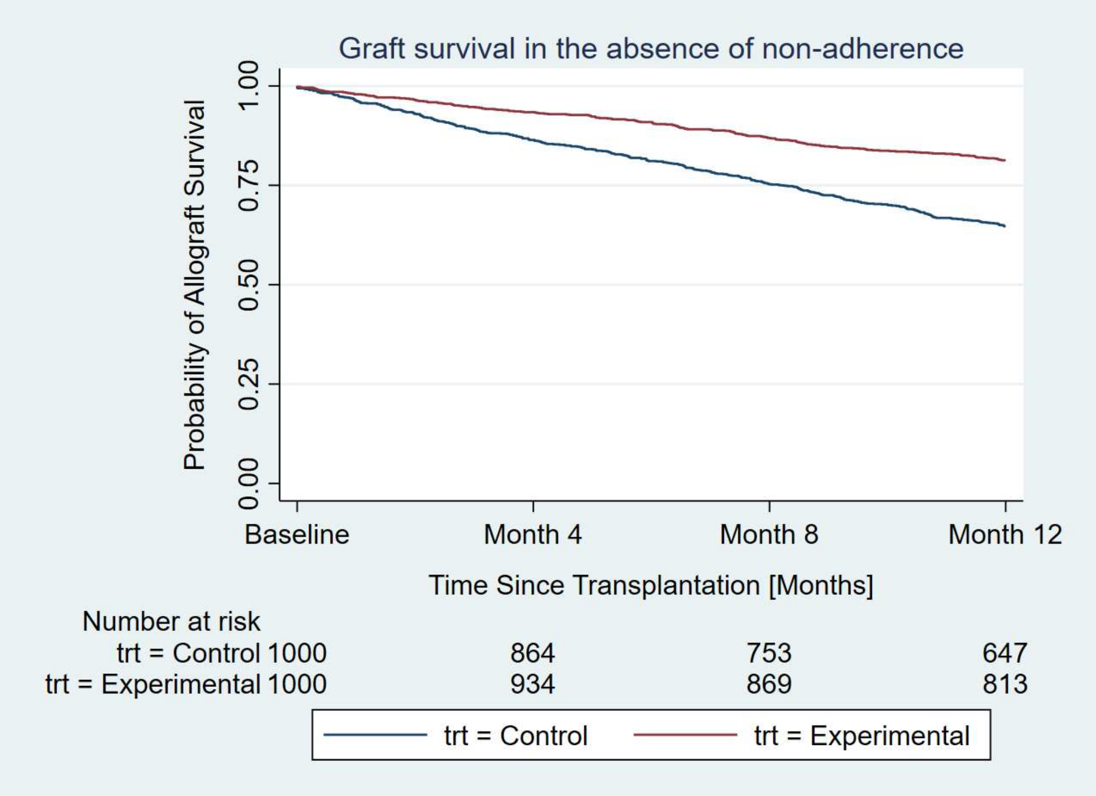
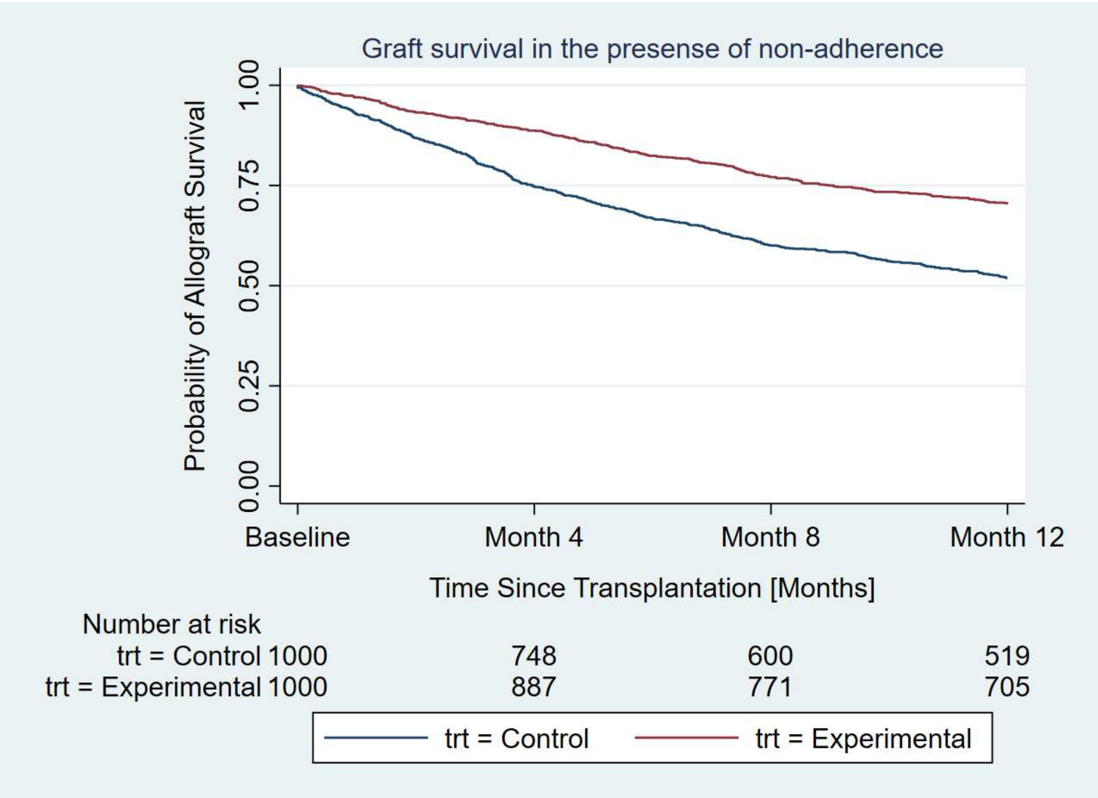


Figure 11: Graft survival curves from the simulated dataset in the presence of non-adherence - Scenario 2



To illustrate the impact of non-adherence and prognostic characteristics on graft survival over time, KM survival curves were generated based on analysis of the above-mentioned dataset. These include the impact of age (Figure 5), BMI (Figures 6-8), and non-adherence (Figures 9-11).

Figure 12: Impact of age on simulated graft survival time

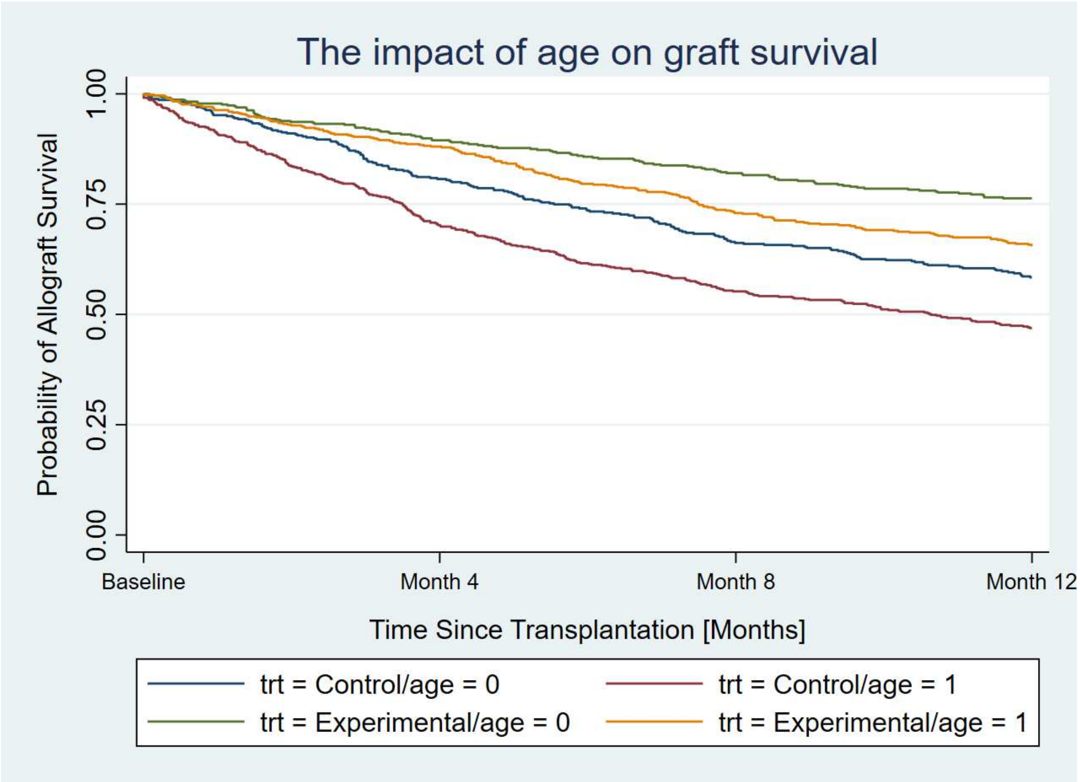


Figure 13: Impact of baseline BMI on simulated graft survival time

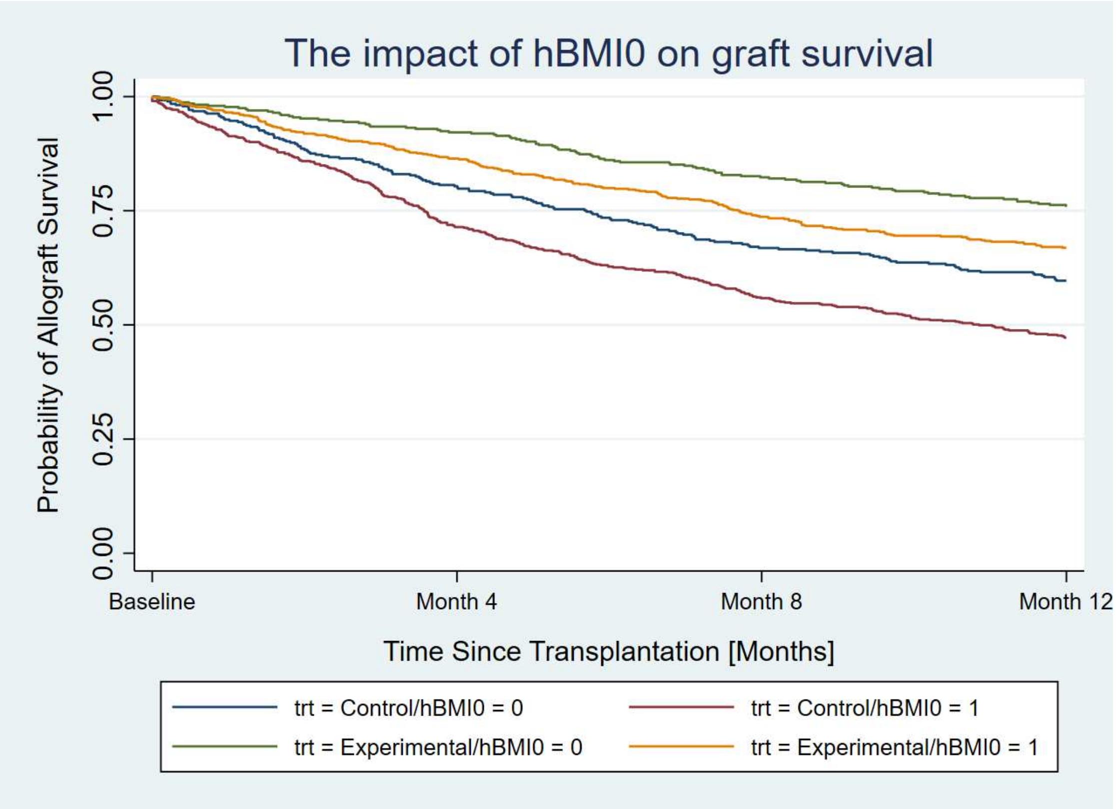


Figure 14: Impact of BMI at 4 months on simulated graft survival time

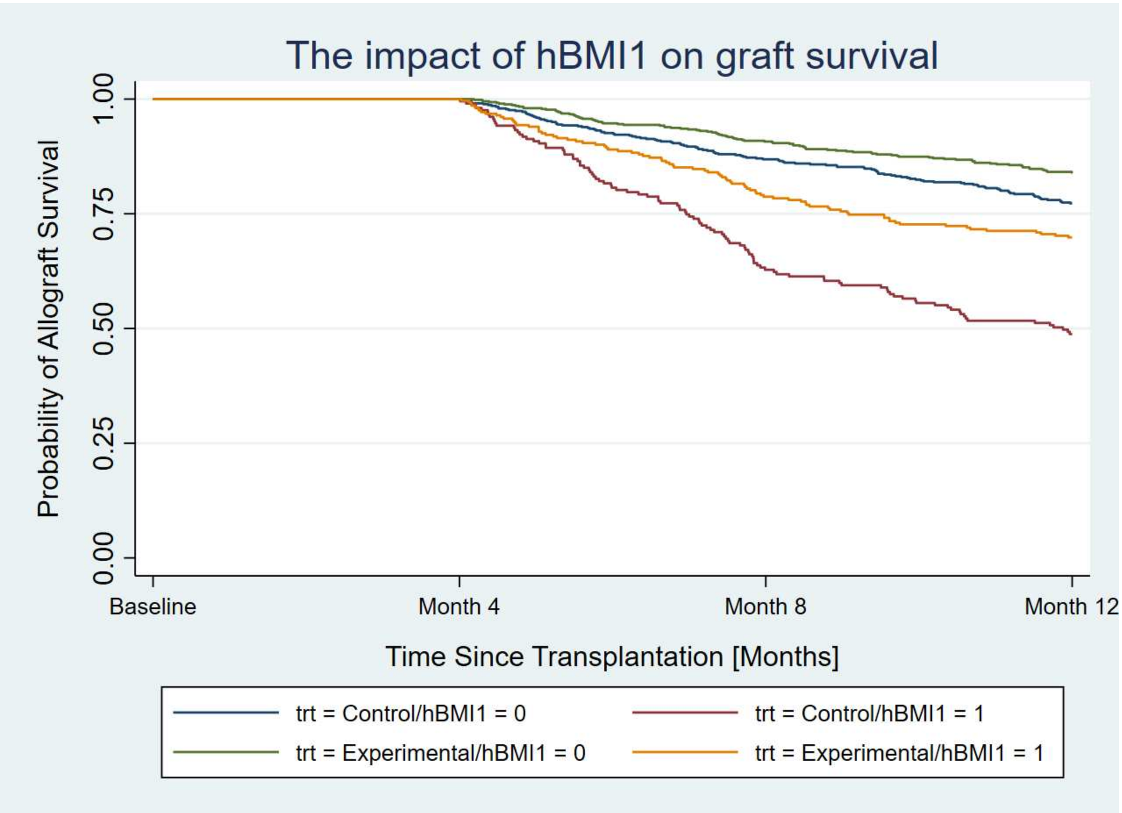


Figure 15: Impact of BMI at 8 months on simulated graft survival time

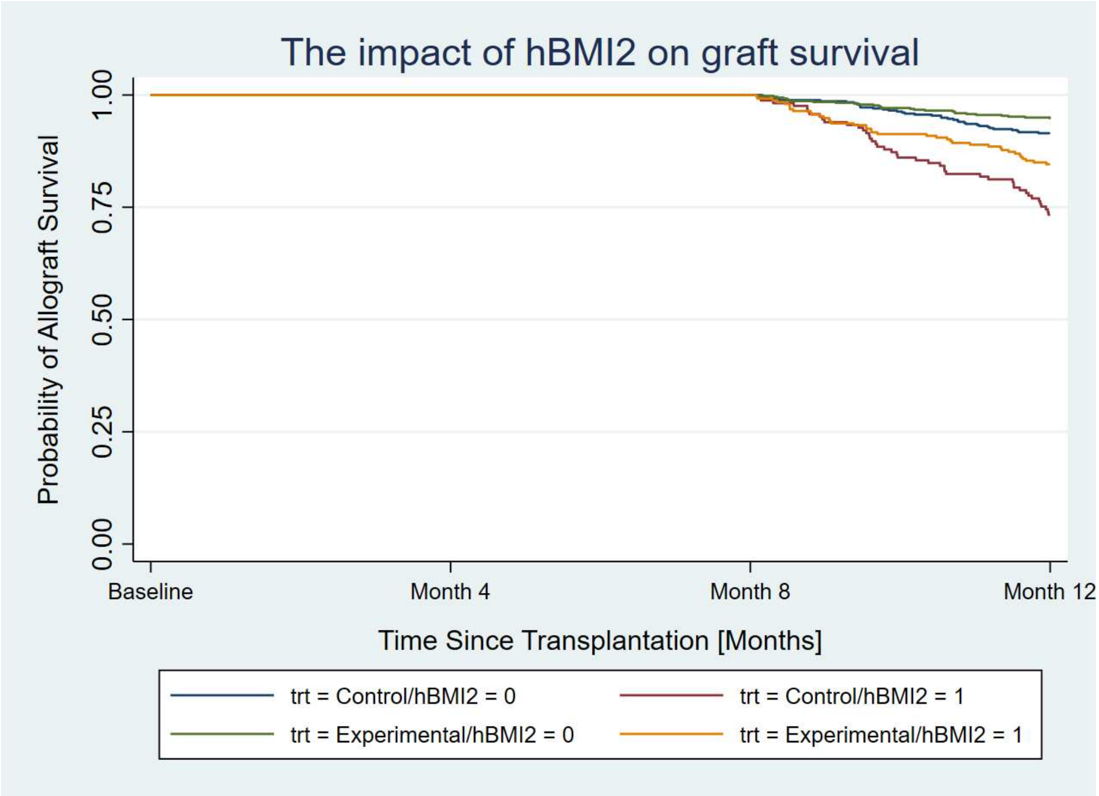


Figure 16: Impact of 0-4 month's implementation non-adherence on graft survival

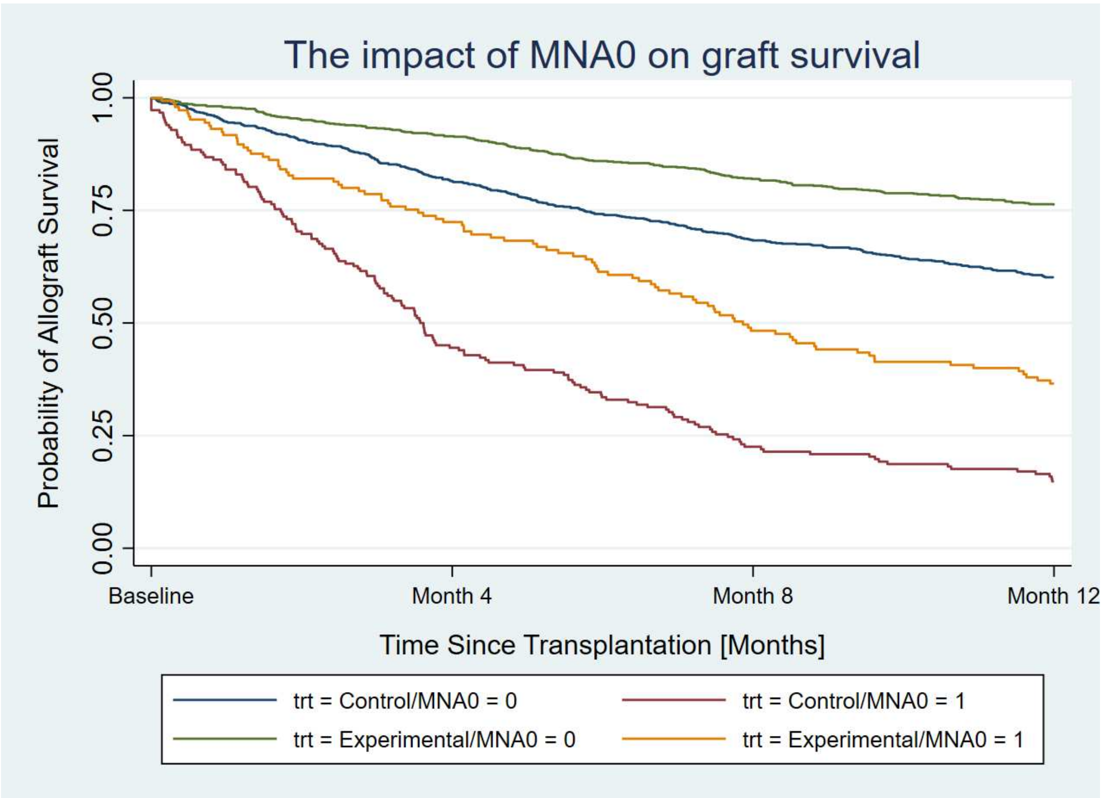


Figure 17: Impact of 4-8 month's implementation non-adherence on graft survival

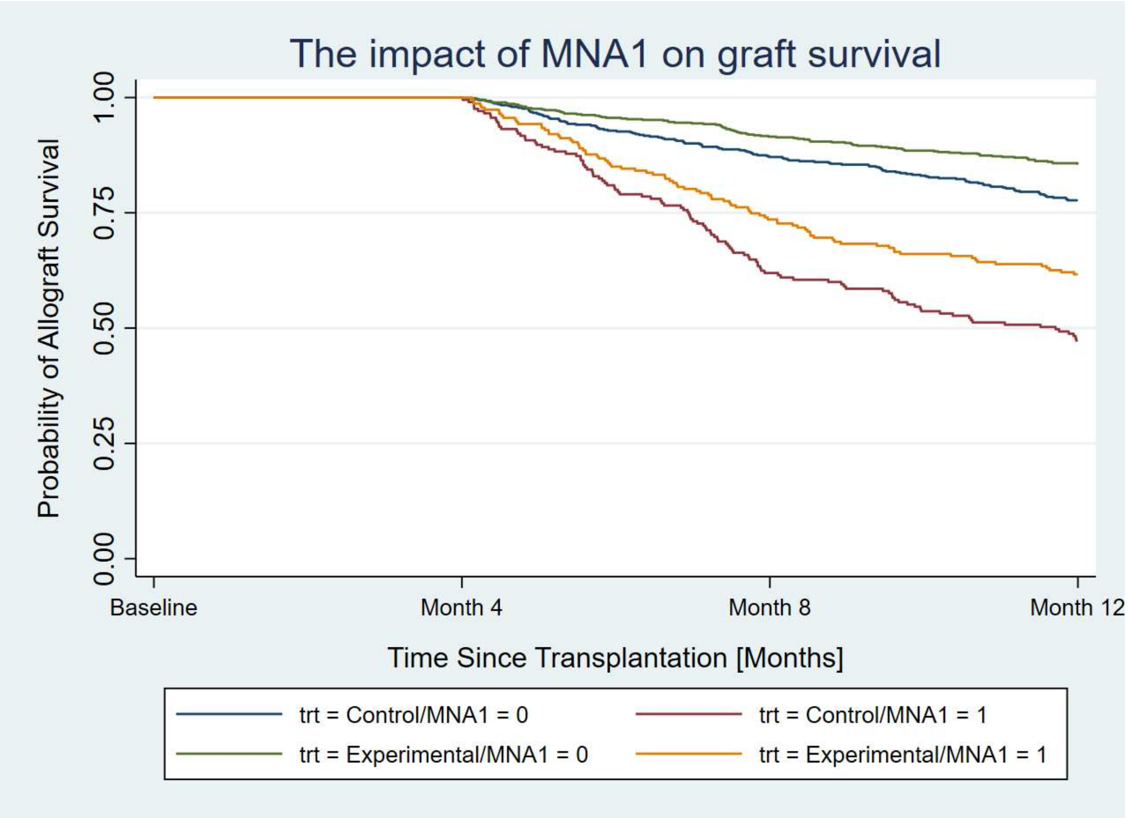
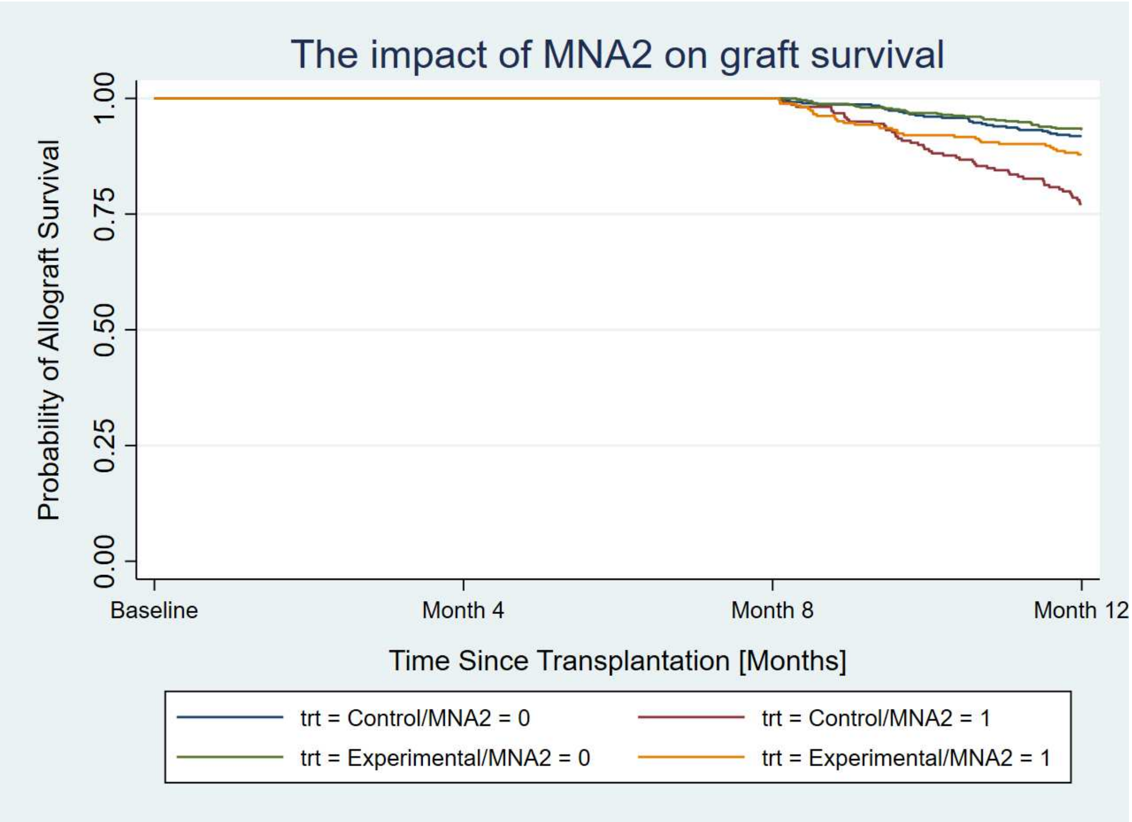


Figure 18: Impact of 8-12 month's implementation non-adherence on graft survival



3. Performance Measures

The performance assessment was focused on the following key properties of the estimators: bias, accuracy, coverage, and both empirical and model-based standard errors. In addition, to quantify simulation uncertainty over n_{sim}, we have estimated the Monte Carlo standard errors (MCSE) of the estimated performance measures. The formulae for computing the performance measures are provided below.

– Bias is a measure of accuracy and was evaluated using absolute and percentage bias. The absolute bias was calculated using the following formula.

$$Bias = \frac{1}{n_{sim}} \sum_{i=1}^{n_{sim}} \hat{\theta}_i - \theta \qquad [1]$$

– Mean Squared Error (MSE) is a measure of overall accuracy because it includes both bias and variability measures. This is presented as a percentage of the true value. Formally, it is the sum of the squared bias and variance of $\hat{\theta}$ which can be computed using the following formula.

$$MSE = \frac{1}{n_{sim}} \sum_{i=1}^{n_{sim}} (\hat{\theta}_i - \theta)^2 \qquad [2]$$

– Coverage is defined as the probability that a CI contains θ (i.e., the proportion of times that 95% CI contains the true value of the estimated parameter). For a two-sided interval, coverage was computed using the following formula.

$$Covergae = \frac{1}{n_{sim}} \sum_{i=1}^{n_{sim}} 1(\hat{\theta}_{low,i} \leq \theta \leq \hat{\theta}_{upp,i}) \qquad [3]$$

- Empirical standard error (EmpSE) for $\hat{\theta}$ is an estimate of the long-run standard deviation of $\hat{\theta}$ over the n_{sim}. The EmpSE was computed using the following formula.

$$EmpSE = \sqrt{\frac{1}{n_{sim} - 1} \sum_{i=1}^{n_{sim}} (\hat{\theta} - \bar{\theta})^2} \qquad [4]$$

- Average model-based standard error (ModSE) for $\hat{\theta}$ is the average of the estimated SEs which targets the estimated empirical standard error. The ModSE was computed using the following formula.

$$Average\ ModSE = \sqrt{\frac{1}{n_{sim} - 1} \sum_{i=1}^{n_{sim}} \widehat{Var}\,(\hat{\theta}_i)} \qquad [5]$$

All performance measures were expressed as a percentage of the true value of the treatment effect (i.e., Difference in RMSTs).

References

1. Jones-Hughes T, Snowsill T, Haasova M, et al. Immunosuppressive therapy for kidney transplantation in adults: a systematic review and economic model. *Health Technol Assess* 2016; 20: 1-594. DOI: 10.3310/hta20620.

2. Pizzo HP, Ettenger RB, Gjertson DW, et al. Sirolimus and tacrolimus coefficient of variation is associated with rejection, donor-specific antibodies, and nonadherence. *Pediatr Nephrol* 2016; 31: 2345–2352. DOI: 10.1007/s00467-016-3422-5.

3. Rodrigo E, Segundo DS, Fernández-Fresnedo G, et al. Within-Patient Variability in Tacrolimus Blood Levels Predicts Kidney Graft Loss and Donor-Specific Antibody Development. *Transplantation* 2016; 100: 2479-2485. DOI: 10.1097/TP.0000000000001040.

4. Crowther MJ and Lambert PC. Simulating biologically plausible complex survival data. *Stat Med* 2013; 32: 4118–4134. DOI: 10.1002/sim.5823.

Supplementary Appendix B: Specification of scenarios assessed in the simulation study

Table 4: Specification of scenarios assessed in the simulation across implementation, persistence and initiation non-adherence

Adjustment Scenario No.	Truth Scenario No.	Sample size	Survival time Data-generating Model (DGM)	Type and level of non-adherence	Relationship between treatment effect and non-adherence	Time-dependent treatment effect	Treatment effect size	Non-adherence adjustment methods assessed
1	1	450	Standard PSM - Weibull	Low implementation	Strong	No	Small	ITT, PP, IPCW, SNFTM
2	2	450	Two-component Weibull Mixture	Low implementation	Strong	No	Small	ITT, PP, IPCW, SNFTM
3	1	450	Standard PSM - Weibull	High implementation	Weak	No	Small	ITT, PP, IPCW, SNFTM
4	2	450	Two-component Weibull Mixture	High implementation	Weak	No	Small	ITT, PP, IPCW, SNFTM
5	1	450	Standard PSM - Weibull	Low implementation	Weak	No	Small	ITT, PP, IPCW, SNFTM
6	2	450	Two-component Weibull Mixture	Low implementation	Weak	No	Small	ITT, PP, IPCW, SNFTM
7	3	450	Standard PSM - Weibull	High implementation	Strong	No	Large	ITT, PP, IPCW, SNFTM
8	4	450	Two-component Weibull Mixture	High implementation	Strong	No	Large	ITT, PP, IPCW, SNFTM
9	3	450	Standard PSM - Weibull	Low implementation	Strong	No	Large	ITT, PP, IPCW, SNFTM
10	4	450	Two-component Weibull Mixture	Low implementation	Strong	No	Large	ITT, PP, IPCW, SNFTM
11	5	450	Standard PSM - Weibull	High implementation	Strong	Yes	Large	ITT, PP, IPCW, SNFTM
12	6	450	Two-component Weibull Mixture	High implementation	Strong	Yes	Large	ITT, PP, IPCW, SNFTM
13	5	450	Standard PSM - Weibull	Low implementation	Strong	Yes	Large	ITT, PP, IPCW, SNFTM
14	6	450	Two-component Weibull Mixture	Low implementation	Strong	Yes	Large	ITT, PP, IPCW, SNFTM
15	7	450	Standard PSM - Weibull	High implementation	Weak	Yes	Small	ITT, PP, IPCW, SNFTM
16	8	450	Two-component Weibull Mixture	High implementation	Weak	Yes	Small	ITT, PP, IPCW, SNFTM
17	7	450	Standard PSM - Weibull	Low implementation	Weak	Yes	Small	ITT, PP, IPCW, SNFTM
18	8	450	Two-component Weibull Mixture	Low implementation	Weak	Yes	Small	ITT, PP, IPCW, SNFTM
19	9	120	Standard PSM - Weibull	High implementation	Strong	No	Small	ITT, PP, IPCW, SNFTM
20	10	120	Two-component Weibull Mixture	High implementation	Strong	No	Small	ITT, PP, IPCW, SNFTM
21	9	120	Standard PSM - Weibull	Low implementation	Strong	No	Small	ITT, PP, IPCW, SNFTM
22	10	120	Two-component Weibull Mixture	Low implementation	Strong	No	Small	ITT, PP, IPCW, SNFTM
23	9	120	Standard PSM - Weibull	High implementation	Weak	No	Small	ITT, PP, IPCW, SNFTM
24	10	120	Two-component Weibull Mixture	High implementation	Weak	No	Small	ITT, PP, IPCW, SNFTM
25	9	120	Standard PSM - Weibull	Low implementation	Weak	No	Small	ITT, PP, IPCW, SNFTM
26	10	120	Two-component Weibull Mixture	Low implementation	Weak	No	Small	ITT, PP, IPCW, SNFTM

Adjustment Scenario No.	Truth Scenario No.	Sample size	Survival time Data-generating Model (DGM)	Type and level of non-adherence	Relationship between treatment effect and non-adherence	Time-dependent treatment effect	Treatment effect size	Non-adherence adjustment methods assessed
27	11	120	Standard PSM - Weibull	High implementation	Strong	No	Large	ITT, PP, IPCW, SNFTM
28	12	120	Two-component Weibull Mixture	High implementation	Strong	No	Large	ITT, PP, IPCW, SNFTM
29	11	120	Standard PSM - Weibull	Low implementation	Strong	No	Large	ITT, PP, IPCW, SNFTM
30	12	120	Two-component Weibull Mixture	Low implementation	Strong	No	Large	ITT, PP, IPCW, SNFTM
31	13	120	Standard PSM - Weibull	High implementation	Strong	Yes	Large	ITT, PP, IPCW, SNFTM
32	14	120	Two-component Weibull Mixture	High implementation	Strong	Yes	Large	ITT, PP, IPCW, SNFTM
33	13	120	Standard PSM - Weibull	Low implementation	Strong	Yes	Large	ITT, PP, IPCW, SNFTM
34	14	120	Two-component Weibull Mixture	Low implementation	Strong	Yes	Large	ITT, PP, IPCW, SNFTM
35	15	120	Standard PSM - Weibull	High implementation	Weak	Yes	Small	ITT, PP, IPCW, SNFTM
36	16	120	Two-component Weibull Mixture	High implementation	Weak	Yes	Small	ITT, PP, IPCW, SNFTM
37	15	120	Standard PSM - Weibull	Low implementation	Weak	Yes	Small	ITT, PP, IPCW, SNFTM
38	16	120	Two-component Weibull Mixture	Low implementation	Weak	Yes	Small	ITT, PP, IPCW, SNFTM
39	1	450	Standard PSM - Weibull	Low persistence	Strong	No	Small	ITT, PP, IPCW, SNFTM
40	2	450	Two-component Weibull Mixture	Low persistence	Strong	No	Small	ITT, PP, IPCW, SNFTM
41	1	450	Standard PSM - Weibull	High persistence	Weak	No	Small	ITT, PP, IPCW, SNFTM
42	2	450	Two-component Weibull Mixture	High persistence	Weak	No	Small	ITT, PP, IPCW, SNFTM
43	4	450	Two-component Weibull Mixture	High persistence	Strong	No	Large	ITT, PP, IPCW, SNFTM
44	3	450	Standard PSM - Weibull	Low persistence	Strong	No	Large	ITT, PP, IPCW, SNFTM
45	4	450	Two-component Weibull Mixture	Low persistence	Strong	No	Large	ITT, PP, IPCW, SNFTM
46	5	450	Standard PSM - Weibull	High persistence	Strong	Yes	Large	ITT, PP, IPCW, SNFTM
47	6	450	Two-component Weibull Mixture	High persistence	Strong	Yes	Large	ITT, PP, IPCW, SNFTM
48	5	450	Standard PSM - Weibull	Low persistence	Strong	Yes	Large	ITT, PP, IPCW, SNFTM
49	6	450	Two-component Weibull Mixture	Low persistence	Strong	Yes	Large	ITT, PP, IPCW, SNFTM
50	10	120	Two-component Weibull Mixture	High persistence	Strong	No	Small	ITT, PP, IPCW, SNFTM
51	12	120	Two-component Weibull Mixture	High persistence	Strong	No	Large	ITT, PP, IPCW, SNFTM
52	12	120	Two-component Weibull Mixture	Low persistence	Strong	No	Large	ITT, PP, IPCW, SNFTM
53	16	120	Two-component Weibull Mixture	High persistence	Weak	Yes	Small	ITT, PP, IPCW, SNFTM
54	15	120	Standard PSM - Weibull	Low persistence	Weak	Yes	Small	ITT, PP, IPCW, SNFTM
55	16	120	Two-component Weibull Mixture	Low persistence	Weak	Yes	Small	ITT, PP, IPCW, SNFTM
56	14	120	Two-component Weibull Mixture	High persistence	Strong	Yes	Large	ITT, PP, IPCW, SNFTM
57	1	450	Standard PSM - Weibull	Low initiation	Strong	No	Small	ITT, PP, IPCW, SNFTM
58	2	450	Two-component Weibull Mixture	Low initiation	Strong	No	Small	ITT, PP, IPCW, SNFTM
59	1	450	Standard PSM - Weibull	High initiation	Weak	No	Small	ITT, PP, IPCW, SNFTM

Adjustment Scenario No.	Truth Scenario No.	Sample size	Survival time Data-generating Model (DGM)	Type and level of non-adherence	Relationship between treatment effect and non-adherence	Time-dependent treatment effect	Treatment effect size	Non-adherence adjustment methods assessed
60	2	450	Two-component Weibull Mixture	High initiation	Weak	No	Small	ITT, PP, IPCW, SNFTM
61	1	450	Standard PSM - Weibull	Low initiation	Weak	No	Small	ITT, PP, IPCW, SNFTM
62	2	450	Two-component Weibull Mixture	Low initiation	Weak	No	Small	ITT, PP, IPCW, SNFTM
63	3	450	Standard PSM - Weibull	High initiation	Strong	No	Large	ITT, PP, IPCW, SNFTM
64	4	450	Two-component Weibull Mixture	High initiation	Strong	No	Large	ITT, PP, IPCW, SNFTM
65	3	450	Standard PSM - Weibull	Low initiation	Strong	No	Large	ITT, PP, IPCW, SNFTM
66	4	450	Two-component Weibull Mixture	Low initiation	Strong	No	Large	ITT, PP, IPCW, SNFTM
67	5	450	Standard PSM - Weibull	High initiation	Strong	Yes	Large	ITT, PP, IPCW, SNFTM
68	6	450	Two-component Weibull Mixture	High initiation	Strong	Yes	Large	ITT, PP, IPCW, SNFTM
69	5	450	Standard PSM - Weibull	Low initiation	Strong	Yes	Large	ITT, PP, IPCW, SNFTM
70	6	450	Two-component Weibull Mixture	Low initiation	Strong	Yes	Large	ITT, PP, IPCW, SNFTM
71	7	450	Standard PSM - Weibull	High initiation	Weak	Yes	Small	ITT, PP, IPCW, SNFTM
72	8	450	Two-component Weibull Mixture	High initiation	Weak	Yes	Small	ITT, PP, IPCW, SNFTM
73	7	450	Standard PSM - Weibull	Low initiation	Weak	Yes	Small	ITT, PP, IPCW, SNFTM
74	8	450	Two-component Weibull Mixture	Low initiation	Weak	Yes	Small	ITT, PP, IPCW, SNFTM
75	9	120	Standard PSM - Weibull	High initiation	Strong	No	Large	ITT, PP, IPCW, SNFTM
76	10	120	Two-component Weibull Mixture	High initiation	Strong	No	Large	ITT, PP, IPCW, SNFTM
77	9	120	Standard PSM - Weibull	Low initiation	Strong	No	Large	ITT, PP, IPCW, SNFTM
78	10	120	Two-component Weibull Mixture	Low initiation	Strong	No	Large	ITT, PP, IPCW, SNFTM
79	9	120	Standard PSM - Weibull	High initiation	Weak	No	Large	ITT, PP, IPCW, SNFTM
80	10	120	Two-component Weibull Mixture	High initiation	Weak	No	Large	ITT, PP, IPCW, SNFTM
81	9	120	Standard PSM - Weibull	Low initiation	Weak	No	Large	ITT, PP, IPCW, SNFTM
82	10	120	Two-component Weibull Mixture	Low initiation	Weak	No	Large	ITT, PP, IPCW, SNFTM
83	13	120	Standard PSM - Weibull	High initiation	Strong	Yes	Small	ITT, PP, IPCW, SNFTM
84	14	120	Two-component Weibull Mixture	High initiation	Strong	Yes	Small	ITT, PP, IPCW, SNFTM
85	13	120	Standard PSM - Weibull	Low initiation	Strong	Yes	Small	ITT, PP, IPCW, SNFTM
86	14	120	Two-component Weibull Mixture	Low initiation	Strong	Yes	Small	ITT, PP, IPCW, SNFTM
87	15	120	Standard PSM - Weibull	High initiation	Weak	Yes	Large	ITT, PP, IPCW, SNFTM
88	16	120	Two-component Weibull Mixture	High initiation	Weak	Yes	Large	ITT, PP, IPCW, SNFTM
89	15	120	Standard PSM - Weibull	Low initiation	Weak	Yes	Large	ITT, PP, IPCW, SNFTM
90	16	120	Two-component Weibull Mixture	Low initiation	Weak	Yes	Large	ITT, PP, IPCW, SNFTM

Note: For each set of scenarios numbered in the first column, there was one large dataset (truth scenario numbered in the second column) which was simulated using 1 million iterations

Supplementary Appendix C: Code used in the simulation program

The code used to run the simulation study in Stata MP (version 15.1) is presented below. The first set of code is for generating the truth for Scenario 1 with factor specifications as follows: sample size (n=450), standard PSM DGM, perfect implementation non-adherence, no time-dependent treatment effect and large treatment effect size. These factors were amended to run the simulation program for each scenario as specified in Supplementary Appendix B. The second set of code incorporate non-adherence and applies the alternative adjustment methods with factors amended across 90 scenarios.

Simulation program to generate the truth

```
*** Install Stata packages***
ssc install randomize
ssc install survsim
ssc install moremata
ssc install rcsgen
ssc install stpm2
ssc install simsum
net from https://www.pclambert.net/downloads/standsurv
install stgest3  // manual installation
install nlplot     // manual installation

*** Truth 1***
clear
capture program drop truth1
program define truth1, rclass
version 15.1
scalar drop _all

*** Run the simulation code to produce RCT datasets with perfect adherence and estimates dataset
quietly {
 local nobs 450    // number of observations in each simulated data set
 set varabbrev off

            set coeftabresults off //  runs faster
  drop _all

  *** declare sample size
  set obs `nobs'

            * Generate baseline covariates
            gen id= _n
            gen random = uniform()
            gen age = cond(random < 0.55, 1, 0)

            * Generate time-varying covariate at baseline
            gen hBMI0 = rbinomial(1,0.60)

            * Randomise observation to two groups "1= experimental and 2= control"
            randomize, groups(2) balance(age hBMI0)
            recode _assignment (1 = 0) (2 = 1), gen(trt)

            ** Generate hBMI at 4 months influenced by MNA0 and hBMI0
            gen hBMI1=rbinomial(1,0.30) if hBMI0==1
            replace hBMI1=rbinomial(1,0.20) if hBMI0==0

            ** Generate hBMI at 8 months influenced by MNA1 and hBMI1
            gen hBMI2=rbinomial(1,0.30) if hBMI1==1
            replace hBMI2=rbinomial(1,0.20) if hBMI1==0
```

*** DGMs for generating survival times using survsim with a user-defined hazard function incorporating both baseline and time-dependent covariates

```
capture: survsim stime event, loghazard(-1.2:*0.2:*#t:^(0.2:-1) :* (hBMI0:* 0.35) :* (0:<=#t:<0.3333333) :+ (hBMI1:* 0.35) :* ///
 (0.3333333:<=#t:<0.6666667) :+ (hBMI2:* 0.35) :* (0.6666667:<=#t:<=1)) cov(trt -0.45 age 0.25) maxt(1)

* Declare the data to be survival data
stset stime, failure(event=1) id(id)

	replace hBMI1=. if stime<= float(0.3333333)
	replace hBMI2=. if stime<= float(0.6666667)

	***Estimate the truth
	capture stcox trt
	if (e(converged)>0) return scalar conv_hr=1
	else return scalar conv_hr= 0
	if (_rc>0) return scalar error_hr=1
	else return scalar error_hr=0
	if (e(converged)>0) {
		return scalar hr=exp(_b[trt])
				return scalar hr_se=exp(_b[trt])*_se[trt]
	}
	else {
		return scalar hr=.
				return scalar hr_se=.
	}

	capture stpm2 trt, scale(h) df(2) lininit nolog eform
	if (e(converged)>0) return scalar conv_rmst=1
	else return scalar conv_rmst= 0
	if (_rc>0) return scalar error_rmst=1
	else return scalar error_rmst=0
	if (e(converged)>0) {
				gen tt1 = 1 in 1
				standsurv, at1(trt 0) at2(trt 1) ci se timevar(tt1) contrast(difference) rmst atvars(rmst_trt0 
rmst_trt1) contrastvar(rmst_diff)
				summ rmst_diff, meanonly
				return scalar rmstdiff= r(mean)
				summ rmst_diff_se, meanonly
				return scalar rmstdiff_se= r(mean)
				summ rmst_diff_lci, meanonly
				return scalar rmstdiff_lci= r(mean)
				summ rmst_diff_uci, meanonly
				return scalar rmstdiff_uci= r(mean)
				summ rmst_trt0, meanonly
				return scalar rmst0= r(mean)
				summ rmst_trt0_se, meanonly
				return scalar rmst_trt0_se=r(mean)
				summ rmst_trt0_lci, meanonly
				return scalar rmst0_lci= r(mean)
				summ rmst_trt0_uci, meanonly
				return scalar rmst0_uci= r(mean)
				summ rmst_trt1, meanonly
				return scalar rmst1= r(mean)
				summ rmst_trt1_se, meanonly
				return scalar rmst_trt1_se=r(mean)
				summ rmst_trt1_lci, meanonly
				return scalar rmst1_lci= r(mean)
				summ rmst_trt1_uci, meanonly
				return scalar rmst1_uci= r(mean)
	}
	else {
				return scalar rmstdiff=.
				return scalar rmstdiff_se=.
				return scalar rmstdiff_lci=.
				return scalar rmstdiff_uci=.
				return scalar rmst0=.
				return scalar rmst_trt0_se=.
				return scalar rmst0_lci=.
				return scalar rmst0_uci=.
				return scalar rmst1=.
				return scalar rmst_trt1_se=.
```

```
                                                return scalar rmst1_lci=.
                                                return scalar rmst1_uci=.
                                }
}
end

set rng mt64s   // set the  stream 64-bit Mersenne Twister
set rngstream 11       //set the stream of rng
simulate hr=r(hr)  ///
hr_se=r(hr_se) ///
conv_hr=r(conv_hr) ///
error_hr=r(error_hr) ///
rmstdiff= r(rmstdiff) ///
rmstdiff_se = r(rmstdiff_se) ///
rmstdiff_lci= r(rmstdiff_lci) ///
rmstdiff_uci= r(rmstdiff_uci) ///
rmst0= r(rmst0) ///
rmst_trt0_se= r(rmst_trt0_se) ///
rmst0_lci= r(rmst0_lci) ///
rmst0_uci= r(rmst0_uci) ///
rmst1= r(rmst1) ///
rmst_trt1_se= r(rmst_trt1_se) ///
rmst1_lci= r(rmst1_lci) ///
rmst1_uci= r(rmst1_uci) ///
conv_rmst=r(conv_rmst) ///
error_rmst=r(error_rmst), ///
reps(1000000) seed(13183) saving(estimates1, replace): truth1

use estimates1, clear
gen idrep= _n   // generate idrep number
order idrep, first
gen dgm= 1   // generate dgm number
order dgm, after(idrep)

***rename variable names to sensible names
rename rmst_trt0_se rmst0_se
rename rmst_trt1_se rmst1_se

*** Order variables in the estimates dataset
order hr, after(rmst1_uci)
order hr_se, after(hr)
order conv_hr, after(error_rmst)
order error_hr, after(conv_hr)

*** Label variables and values
label variable idrep "Rep num"
label variable dgm "Data-generating mechanism"
label variable rmstdiff "Difference in Restricted Mean Survival Times"
label variable rmstdiff_se "Standard Error of the Difference in Restricted Mean Survival Times"
label variable rmstdiff_lci "RMST 95% CI: Upper bound"
label variable rmstdiff_uci "RMST 95% CI: Lower bound"
label variable rmst0 "Restricted Mean Survival Time 'Control Group'"
label variable rmst0_se "Standard Error of RMST 'Control Group'"
label variable rmst0_lci "RMST 95% CI: Lower bound 'Control Group'"
label variable rmst0_uci "RMST 95% CI: Upper bound 'Control Group'"
label variable rmst1 "Restricted Mean Survival Time 'Exp Group'"
label variable rmst1_se "Standard Error of RMST 'Exp Group'"
label variable rmst1_lci "RMST 95% CI: Lower bound 'Exp Group'"
label variable rmst1_uci "RMST 95% CI: Upper bound 'Exp Group'"
label variable hr "Hazard Ratio"
label variable hr_se "Standard Error of Hazard Ratio"
label variable conv_hr "HR model converged"
label variable error_hr "Error - HR model"
label variable conv_rmst "RMST model converged"
label variable error_rmst "Error - RMST model"
label define nylab 0 "No" 1 "Yes"
label values conv_hr conv_rmst error_hr error_rmst nylab
label define dgmlab 1 "Standard PSM - Weibull" 2 "Two-component Weibull Mixture"
label values dgm dgmlab

*** Save the truth estimates dataset
```

```
save truth1, replace
```

Simulation program to apply non-adherence adjustment methods

```
*** Non-Adherence Adjustment: Scenario 1 ***
clear
capture program drop mysimc1
program define mysimc1, rclass
version 15.1
scalar drop _all

**Run the simulation to produce non-adherence adjusted estimates dataset
quietly {
 local nobs 450                    // number of observations in each simulated data set
 set varabbrev off

          set coeftabresults off //  runs faster
  drop _all

  * declare sample size
  set obs `nobs'

          * Generate baseline covariates
          gen id= _n
          gen random = uniform()
          gen age = cond(random < 0.55, 1, 0)

          * Generate time-varying covariate at baseline
          gen hBMI0 = rbinomial(1,0.60)

          * Randomise observation to two groups "1= experimental and 2= control"
          randomize, groups(2) balance(age hBMI0)
          recode _assignment (1 = 0) (2 = 1), gen(trt)

          * MNA0: Generate time-varying non-adherence (implementation) and covariates at follow-up time 1 (Month 4 for tdc, 0-4 Months for MNA)
          *** Exp Group
          *** This assumes people with high BMI and age <24 have 22.5% chance of MNA.
          *** This risk is reduced to 15% if they have high BMI but age> 24 and 10% risk if age<24 but normal BMI
          *** For people with normal BMI and age>24, the probability of MNA is 5%
          gen MNA0=rbinomial(1,0.225) if (age==1 & hBMI0==1 & trt==1)
          replace MNA0=rbinomial(1,0.15) if (age==0 & hBMI0==1 & trt==1)
          replace MNA0=rbinomial(1,0.10) if (age==1 & hBMI0==0 & trt==1)
          replace MNA0=rbinomial(1,0.05) if (age==0 & hBMI0==0 & trt==1)

          *** Control Group
          *** This assumes people with high BMI and age <24 have 30% chance of MNA (higher than in the control group).
          *** This risk is reduced to 20% if they have high BMI but age> 24 and 10% risk if age<24 but normal BMI
          *** For people with normal BMI and age>24, the probability of MNA is 5%
          replace MNA0=rbinomial(1,0.30) if (age==1 & hBMI0==1 & trt==0)
          replace MNA0=rbinomial(1,0.20) if (age==0 & hBMI0==1 & trt==0)
          replace MNA0=rbinomial(1,0.10) if (age==1 & hBMI0==0 & trt==0)
          replace MNA0=rbinomial(1,0.05) if (age==0 & hBMI0==0 & trt==0)

          ** Generate hBMI at 4 months influenced by MNA0 and hBMI0
          *** This assumes people high BMI at baseline and non-adhered between 0-4 months will have 90% chance to have high BMI at Month 4 and this risk is reduced to 30% among adhered
          *** People with normal BMI at baseline with MNA0=1 have 60% chance of moving to high BMI category and this risk is reduced to 20% among adhered people
          *** The strength of the relationship between previous hBMI/MNA and subsequent hBMI is assumed to be constant over time
          *** The strgenth of hBMI/MNA relationships is assumed to be similar between the two arms
          gen hBMI1=rbinomial(1,0.90) if (MNA0==1 & hBMI0==1)
          replace hBMI1=rbinomial(1,0.30) if (MNA0==0 & hBMI0==1)
          replace hBMI1=rbinomial(1,0.60) if (MNA0==1 & hBMI0==0)
          replace hBMI1=rbinomial(1,0.20) if (MNA0==0 & hBMI0==0)


          * MNA1: Generate time-varying non-adherence and covariates at follow-up time 2 (Month 8 for tdc, 4-8 Months for MNA)
          *** EXP GROUP
```

```
*** This assumes people with high BMI and age <24 have 45% chance of MNA.
*** This risk is reduced to 30% if they have high BMI but age> 24 and 15% risk if age<24 but normal BMI
*** For people with normal BMI and age>24, the probability of MNA is 7.5%
gen MNA1=rbinomial(1,0.45) if (age==1 & hBMI1==1 & trt==1)
replace MNA1=rbinomial(1,0.30) if (age==0 & hBMI1==1 & trt==1)
replace MNA1=rbinomial(1,0.15) if (age==1 & hBMI1==0 & trt==1)
replace MNA1=rbinomial(1,0.075) if (age==0 & hBMI1==0 & trt==1)

*** Control Group
*** This assumes people with high BMI and age <24 have 60% chance of MNA.
*** This risk is reduced to 40% if they have high BMI but age> 24 and 20% risk if age<24 but normal BMI
*** For people with normal BMI and age>24, the probability of MNA is 10%
replace MNA1=rbinomial(1,0.60) if (age==1 & hBMI1==1 & trt==0)
replace MNA1=rbinomial(1,0.40) if (age==0 & hBMI1==1 & trt==0)
replace MNA1=rbinomial(1,0.20) if (age==1 & hBMI1==0 & trt==0)
replace MNA1=rbinomial(1,0.10) if (age==0 & hBMI1==0 & trt==0)
replace MNA1=1 if MNA0==1

** Generate hBMI at 8 months influenced by MNA1 and hBMI1
** Similar probabilities as hBMI1 at 4 months
gen hBMI2=rbinomial(1,0.90) if (MNA1==1 & hBMI1==1)
replace hBMI2=rbinomial(1,0.30) if (MNA1==0 & hBMI1==1)
replace hBMI2=rbinomial(1,0.60) if (MNA1==1 & hBMI1==0)
replace hBMI2=rbinomial(1,0.20) if (MNA1==0 & hBMI1==0)

* MNA2: Generate time-varying non-adherence at follow-up time 3 (8-12 Months)
*** EXP GROUP
*** This assumes people with high BMI and age <24 have 45% chance of MNA.
*** This risk is reduced to 30% if they have high BMI but age> 24 and 15% risk if age<24 but normal BMI
*** For people with normal BMI and age>24, the probability of MNA is 7.5%
gen MNA2=rbinomial(1,0.45) if (age==1 & hBMI2==1 & trt==1)
replace MNA2=rbinomial(1,0.30) if (age==0 & hBMI2==1 & trt==1)
replace MNA2=rbinomial(1,0.15) if (age==1 & hBMI2==0 & trt==1)
replace MNA2=rbinomial(1,0.075) if (age==0 & hBMI2==0 & trt==1)

*** Control Group
*** This assumes people with high BMI and age <24 have 60% chance of MNA
*** This risk is reduced to 40% if they have high BMI but age> 24 and 20% risk if age<24 but normal BMI
*** For people with normal BMI and age>24, the probability of MNA is 10%
replace MNA2=rbinomial(1,0.60) if (age==1 & hBMI2==1 & trt==0)
replace MNA2=rbinomial(1,0.40) if (age==0 & hBMI2==1 & trt==0)
replace MNA2=rbinomial(1,0.20) if (age==1 & hBMI2==0 & trt==0)
replace MNA2=rbinomial(1,0.10) if (age==0 & hBMI2==0 & trt==0)
replace MNA2=1 if MNA1==1

*** Generate admin censoring time at 1 year (End of study follow-up)
gen admin = 1

* DGMs for generating survival times using survsim with a user-defined hazard function incorporating both baseline and time-dependent covariates
* The model use coefficient values of 0.35 for BMI and 0.40 for non-adherence
        capture: survsim stime event, loghazard(-1.2:*0.2:*#t:^(0.2:-1) :* (hBMI0:* 0.35) :* (0:<=#t:<0.3333333) :+  ///
         (MNA0:* 0.40) :* (0:<=#t:<0.3333333) :+ (hBMI1:* 0.35) :* (0.3333333:<=#t:<0.6666667) :+ ///
        (MNA1:* 0.40) :* (0.3333333:<=#t:<0.6666667) :+ (hBMI2:* 0.35) :* (0.6666667:<=#t:<=1) :+ ///
        (MNA2:* 0.40) :* (0.6666667:<=#t:<=1)) cov(trt -0.45 age 0.25) maxt(1)

*** Declare the data to be survival data
stset stime, failure(event=1) id(id)

        replace hBMI1=. if stime<= float(0.3333333)
        replace hBMI2=. if stime<= float(0.6666667)

        replace MNA2 = . if stime<0.6666667
        replace MNA1 = . if stime<0.3333333

        ****************************************
        * Method 1: ITT
        ****************************************
        preserve
        capture stcox trt age
        if (e(converged)>0) return scalar conv_hr_method1=1
```

```
else return scalar conv_hr_method1= 0
if (_rc>0) return scalar error_hr_method1=1
else return scalar error_hr_method1=0

if (e(converged)>0) {
	return scalar hr_method1=exp(_b[trt])
				return scalar hr_se_method1=exp(_b[trt])*_se[trt]
}
else {
	return scalar hr_method1=.
				return scalar hr_se_method1=.
}

capture stpm2 trt age, scale(h) df(2) lininit nolog eform vce(robust)
if (e(converged)>0) return scalar conv_rmst_method1=1
else return scalar conv_rmst_method1= 0
if (_rc>0) return scalar error_rmst_method1=1
else return scalar error_rmst_method1=0

if (e(converged)>0) {
			gen tt1 = 1 in 1
			standsurv, at1(trt 0) at2(trt 1) ci se timevar(tt1) contrast(difference) rmst atvars(rmst_trt0
rmst_trt1) contrastvar(rmst_diff)
			summ rmst_diff, meanonly
			return scalar rmstdiff_method1= r(mean)
			summ rmst_diff_se, meanonly
			return scalar rmstdiff_se_method1= r(mean)
			summ rmst_diff_lci, meanonly
			return scalar rmstdiff_lci_method1= r(mean)
			summ rmst_diff_uci, meanonly
			return scalar rmstdiff_uci_method1= r(mean)
			summ rmst_trt0, meanonly
			return scalar rmst0_method1= r(mean)
			summ rmst_trt0_se, meanonly
			return scalar rmst_trt0_se_method1=r(mean)
			summ rmst_trt0_lci, meanonly
			return scalar rmst0_lci_method1= r(mean)
			summ rmst_trt0_uci, meanonly
			return scalar rmst0_uci_method1= r(mean)
			summ rmst_trt1, meanonly
			return scalar rmst1_method1= r(mean)
			summ rmst_trt1_se, meanonly
			return scalar rmst_trt1_se_method1=r(mean)
			summ rmst_trt1_lci, meanonly
			return scalar rmst1_lci_method1= r(mean)
			summ rmst_trt1_uci, meanonly
			return scalar rmst1_uci_method1= r(mean)
}
else {
			return scalar rmstdiff_method1=.
			return scalar rmstdiff_se_method1=.
			return scalar rmstdiff_lci_method1=.
			return scalar rmstdiff_uci_method1=.
			return scalar rmst0_method1=.
			return scalar rmst_trt0_se_method1=.
			return scalar rmst0_lci_method1=.
			return scalar rmst0_uci_method1=.
			return scalar rmst1_method1=.
			return scalar rmst_trt1_se_method1=.
			return scalar rmst1_lci_method1=.
			return scalar rmst1_uci_method1=.
}

****************************************
* Method 2: PP
****************************************
restore, preserve
gen eventPP = event
gen stimePP = stime
gen ctime=.
replace ctime= float(0.6666667) if MNA2==1
```

```
replace eventPP=0 if MNA2==1
replace stimePP=ctime if MNA2==1
replace ctime= float(0.3333333) if MNA1==1
replace eventPP=0 if MNA1==1
replace stimePP=ctime if MNA1==1
replace ctime= float(0.0000001) if MNA0==1
replace eventPP=0 if MNA0==1
replace stimePP=ctime if MNA0==1
stset stimePP, failure(eventPP) id(id)

capture stcox trt age
if (e(converged)>0) return scalar conv_hr_method2=1
else return scalar conv_hr_method2= 0
if (_rc>0) return scalar error_hr_method2=1
else return scalar error_hr_method2=0

if (e(converged)>0) {
	return scalar hr_method2=exp(_b[trt])
			return scalar hr_se_method2=exp(_b[trt])*_se[trt]
}
else {
	return scalar hr_method2=.
			return scalar hr_se_method2=.
}

capture stpm2 trt age, scale(h) df(2) lininit nolog eform vce(robust)
if (e(converged)>0) return scalar conv_rmst_method2=1
else return scalar conv_rmst_method2= 0
if (_rc>0) return scalar error_rmst_method2=1
else return scalar error_rmst_method2=0

if (e(converged)>0) {
			gen tt1 = 1 in 1
			standsurv, at1(trt 0) at2(trt 1) ci se timevar(tt1) contrast(difference) rmst atvars(rmst_trt0 
rmst_trt1) contrastvar(rmst_diff)
			summ rmst_diff, meanonly
			return scalar rmstdiff_method2= r(mean)
			summ rmst_diff_se, meanonly
			return scalar rmstdiff_se_method2= r(mean)
			summ rmst_diff_lci, meanonly
			return scalar rmstdiff_lci_method2= r(mean)
			summ rmst_diff_uci, meanonly
			return scalar rmstdiff_uci_method2= r(mean)
			summ rmst_trt0, meanonly
			return scalar rmst0_method2= r(mean)
			summ rmst_trt0_se, meanonly
			return scalar rmst_trt0_se_method2=r(mean)
			summ rmst_trt0_lci, meanonly
			return scalar rmst0_lci_method2= r(mean)
			summ rmst_trt0_uci, meanonly
			return scalar rmst0_uci_method2= r(mean)
			summ rmst_trt1, meanonly
			return scalar rmst1_method2= r(mean)
			summ rmst_trt1_se, meanonly
			return scalar rmst_trt1_se_method2=r(mean)
			summ rmst_trt1_lci, meanonly
			return scalar rmst1_lci_method2= r(mean)
			summ rmst_trt1_uci, meanonly
			return scalar rmst1_uci_method2= r(mean)
}
else {
			return scalar rmstdiff_method2=.
			return scalar rmstdiff_se_method2=.
			return scalar rmstdiff_lci_method2=.
			return scalar rmstdiff_uci_method2=.
			return scalar rmst0_method2=.
			return scalar rmst_trt0_se_method2=.
			return scalar rmst0_lci_method2=.
			return scalar rmst0_uci_method2=.
			return scalar rmst1_method2=.
			return scalar rmst_trt1_se_method2=.
```

```
			return scalar rmst1_lci_method2=.
			return scalar rmst1_uci_method2=.
}

****************************************
* Method 3: IPCW
*****************************************
restore, preserve

*Reshape data from wide format to long format
reshape long hBMI MNA, i(id) j(time)

*** Create time-dependent non-adherence for this interval
gen visit=time
replace time=float(0.3333333) if time==1
replace time=float(0.6666667) if time==2
drop if time>stime
gen timeend = min(stime,time+0.3333333)

*IPCW Step 1: Censor observations for non-adherence and reformat the data
*********************************************************************

* Generate Non-adherence indicator and time of non-adherence (in years)
gen xoind=MNA
gen xotime=.
replace xotime= 0 if xoind==1 & visit==0
replace xotime= float(0.3333333) if xoind==1 & visit==1
replace xotime= float(0.6666667) if xoind==1 & visit==2
by id: egen xotime1 = min(xotime)
replace xotime=xotime1
drop xotime1

gen xotdo=0
replace xotdo= 1 if (xotime>=time) & (xotime<time+float(0.3333333)) & (xoind==1)
replace xotdo= . if (xotime<time) & (xoind==1)

*** Create time-dependent outcome for graft loss (txlosstdo) in each interval (0-4 months, 4-8 months, 8-12 months)
gen txlosstdo = 0
replace txlosstdo = 1 if event==1 & timeend==float(stime)

*** Stset the data
stset timeend, time0(time) failure(txlosstdo)

* Create dummies for being the first and last observation per patient
by id: gen firstobs = _n==1
by id: gen lastobs  = _n==_N

***Note, things change over time as case mix of patients changes.
gen t1 = 0
replace t1=1 if visit==0
gen t2 = 0
replace t2=1 if visit==1
gen t3 = 0
replace t3=1 if visit==2

***Note that the impact of hBMI depends on age, and vice versa. So need interactions
by id: gen cat1 = 0
by id: replace cat1 = 1 if (age==0 & hBMI==0)
by id: gen cat2 = 0
by id: replace cat2 = 1 if (age==1 & hBMI==0)
by id: gen cat3 = 0
by id: replace cat3 = 1 if (age==0 & hBMI==1)
by id: gen cat4 = 0
by id: replace cat4 = 1 if (age==1 & hBMI==1)

***Then make cat time interaction
by id: gen cat1t1 = cat1*t1
by id: gen cat2t1 = cat2*t1
by id: gen cat3t1 = cat3*t1
by id: gen cat4t1 = cat4*t1
```

```
by id: gen cat1t2 = cat1*t2
by id: gen cat2t2 = cat2*t2
by id: gen cat3t2 = cat3*t2
by id: gen cat4t2 = cat4*t2

by id: gen cat1t3 = cat1*t3
by id: gen cat2t3 = cat2*t3
by id: gen cat3t3 = cat3*t3
by id: gen cat4t3 = cat4*t3

*IPCW Step 2: Model the probability of being censored over time
 ***********************************************************************
*Use logistic regression to predict non-adherence given baseline covariates in the control arm (Non-adherence Model 1) &
*Estimate the probability of non-adherence for each patient-observation included in the regression
logistic xotdo age time if trt==0
if (e(converged)>0) return scalar conv_mna1_method3=1
else return scalar conv_mna1_method3= 0
if (_rc>0) return scalar error_mna1_method3=1
else return scalar error_mna1_method3=0
if (e(converged)>0) {
predict pn_mna if e(sample), pr
}

*Use logistic regression to predict non-adherence given the interaction between baseline and time-updated covariates in the control arm (Non-adherence Model 2) &
*Estimate the probability of non-adherence for each patient-observation included in the regression
logistic xotdo cat1t1 cat2t1 cat3t1 cat4t1 cat1t2 cat2t2 cat3t2 cat4t2 cat1t3 cat2t3 cat3t3 cat4t3 time if trt==0
if (e(converged)>0) return scalar conv_mna2_method3=1
else return scalar conv_mna2_method3= 0
if (_rc>0) return scalar error_mna2_method3=1
else return scalar error_mna2_method3=0
if (e(converged)>0) {
predict pd_mna if e(sample), pr
}

*Use logistic regression to predict non-adherence given baseline covariates in the experimental arm (Non-adherence Model 3)
*Estimate the probability of non-adherence for each patient-observation included in the regression
logistic xotdo age time if trt==1
if (e(converged)>0) return scalar conv_mna3_method3=1
else return scalar conv_mna3_method3= 0
if (_rc>0) return scalar error_mna3_method3=1
else return scalar error_mna3_method3=0
if (e(converged)>0) {
predict pn1_mna if e(sample), pr
}

*Use logistic regression to predict non-adherence given the interaction between baseline and time-updated covariates in the experimental arm (Non-adherence Model 4)
*Estimate the probability of non-adherence for each patient-observation included in the regression
logistic xotdo cat1t1 cat2t1 cat3t1 cat4t1 cat1t2 cat2t2 cat3t2 cat4t2 cat1t3 cat2t3 cat3t3 cat4t3 time if trt==1
if (e(converged)>0) return scalar conv_mna4_method3=1
else return scalar conv_mna4_method3= 0
if (_rc>0) return scalar error_mna4_method3=1
else return scalar error_mna4_method3=0
if (e(converged)>0) {
predict pd1_mna if e(sample), pr
}

replace pn_mna = pn1_mna if trt==1
replace pd_mna = pd1_mna if trt==1
drop pn1_mna pd1_mna

*IPCW Step 3: For each individual at each time, compute the inverse probability of remaining uncensored

***********************************************************************************************************
*Estimate the probabilities of remaining uncensored 'adhered' and the IPCW weights
sort id time
gen num = 1-pn_mna if firstobs
replace num = num[_n-1] * (1-pn_mna) if !firstobs
```

```
gen denom = 1-pd_mna if firstobs
replace denom = num[_n-1] * (1-pd_mna) if !firstobs

gen weight = 1/denom
gen sweight = num/denom

**Decalre the data as survival data with stablised weight incorporated and time0 specified for clustering
stset timeend txlosstdo if xotdo==0 [iw=sweight], time0(time)

*IPCW Step 4: Obtain IPCW RMST & HR estimates
*****************************************************************************************************
** Obtain HR using Cox model
capture stcox trt age
if (e(converged)>0) return scalar conv_hr_method3=1
else return scalar conv_hr_method3= 0
if (_rc>0) return scalar error_hr_method3=1
else return scalar error_hr_method3=0

if (e(converged)>0) {
	return scalar hr_method3=exp(_b[trt])
			return scalar hr_se_method3=exp(_b[trt])*_se[trt]
}
else {
	return scalar hr_method3=.
			return scalar hr_se_method3=.
}

** Obtain IPCW RMST estimates using stpm2 model
capture stpm2 trt age, scale(h) df(2) lininit nolog eform vce(robust)
if (e(converged)>0) return scalar conv_rmst_method3=1
else return scalar conv_rmst_method3= 0
if (_rc>0) return scalar error_rmst_method3=1
else return scalar error_rmst_method3=0

if (e(converged)>0) {
			gen tt1 = 1 in 1
			standsurv, at1(trt 0) at2(trt 1) ci se timevar(tt1) contrast(difference) rmst atvars(rmst_trt0 
rmst_trt1) contrastvar(rmst_diff)
			summ rmst_diff, meanonly
			return scalar rmstdiff_method3= r(mean)
			summ rmst_diff_se, meanonly
			return scalar rmstdiff_se_method3= r(mean)
			summ rmst_diff_lci, meanonly
			return scalar rmstdiff_lci_method3= r(mean)
			summ rmst_diff_uci, meanonly
			return scalar rmstdiff_uci_method3= r(mean)
			summ rmst_trt0, meanonly
			return scalar rmst0_method3= r(mean)
			summ rmst_trt0_se, meanonly
			return scalar rmst_trt0_se_method3=r(mean)
			summ rmst_trt0_lci, meanonly
			return scalar rmst0_lci_method3= r(mean)
			summ rmst_trt0_uci, meanonly
			return scalar rmst0_uci_method3= r(mean)
			summ rmst_trt1, meanonly
			return scalar rmst1_method3= r(mean)
			summ rmst_trt1_se, meanonly
			return scalar rmst_trt1_se_method3=r(mean)
			summ rmst_trt1_lci, meanonly
			return scalar rmst1_lci_method3= r(mean)
			summ rmst_trt1_uci, meanonly
			return scalar rmst1_uci_method3= r(mean)
}
else {
			return scalar rmstdiff_method3=.
			return scalar rmstdiff_se_method3=.
			return scalar rmstdiff_lci_method3=.
			return scalar rmstdiff_uci_method3=.
			return scalar rmst0_method3=.
```

```
			return scalar rmst_trt0_se_method3=.
			return scalar rmst0_lci_method3=.
			return scalar rmst0_uci_method3=.
			return scalar rmst1_method3=.
			return scalar rmst_trt1_se_method3=.
			return scalar rmst1_lci_method3=.
			return scalar rmst1_uci_method3=.
}

************************************
*Method 4: SNFTM with G-estimation
************************************
restore
*** Reshape data from wide format to long format
reshape long hBMI MNA, i(id) j(time)

gen visit=time
replace time=float(0.3333333) if time==1
replace time=float(0.6666667) if time==2
drop if time>stime
gen timeend = min(stime,time+0.3333333)

*** Create vist variable to go into the stgest3 model
replace visit=3 if visit==2
replace visit=2 if visit==1
replace visit=1 if visit==0

*** Generate Non-adherence indicator and time of first non-adherence event
gen xoind=MNA
gen xotime=.
replace xotime= 0 if xoind==1 & visit==1
replace xotime= float(0.3333333) if xoind==1 & visit==2
replace xotime= float(0.6666667) if xoind==1 & visit==3
by id: egen xotime1 = min(xotime)
replace xotime=xotime1
drop xotime1

*** Create time-dependent outcome for graft loss (txlosstdo) in each interval (0-4 months, 4-8 months, 8-12 months)
gen txlosstdo = 0
replace txlosstdo = 1 if event==1 & timeend==float(stime)

*** stset the data
stset timeend, failure(txlosstdo) id(id)

by id: gen adlag = xoind[_n-1]
replace adlag=0 if adlag==.

*** Estimate the Acceleration Factor as the effect of time-dependent non-adherence (MNA) on survival time
*** This should be done for each arm separately using the interaction between baseline and time-dependent
covariates

***Note, things change over time as casemix of patients changes.
gen t1 = 0
replace t1=1 if visit==0
gen t2 = 0
replace t2=1 if visit==1
gen t3 = 0
replace t3=1 if visit==2

*** Note that the impact of hBMI depends on age, and vice versa. So need interaction
by id: gen cat1 = 0
by id: replace cat1 = 1 if (age==0 & hBMI==0)
by id: gen cat2 = 0
by id: replace cat2 = 1 if (age==1 & hBMI==0)
by id: gen cat3 = 0
by id: replace cat3 = 1 if (age==0 & hBMI==1)
by id: gen cat4 = 0
by id: replace cat4 = 1 if (age==1 & hBMI==1)

*** Then make cat time interaction
by id: gen cat1t1 = cat1*t1
```

```
by id: gen cat2t1 = cat2*t1
by id: gen cat3t1 = cat3*t1
by id: gen cat4t1 = cat4*t1

by id: gen cat1t2 = cat1*t2
by id: gen cat2t2 = cat2*t2
by id: gen cat3t2 = cat3*t2
by id: gen cat4t2 = cat4*t2

by id: gen cat1t3 = cat1*t3
by id: gen cat2t3 = cat2*t3
by id: gen cat3t3 = cat3*t3
by id: gen cat4t3 = cat4*t3

*** Estimate Accelaration Factor as the effect of time-dependent non-aderehcne on survivial time
*** This should be done for each arm seperatley and calulate admin censoring
preserve
drop if trt==0

***Run the g-estimation on the Exp group
capture stgest3 xoind cat1t1 cat2t1 cat3t1 cat4t1 cat1t2 cat2t2 cat3t2 cat4t2 cat1t3 cat2t3 cat3t3 cat4t3 adlag, visit(visit) lasttime(admin) range (-5 5) model(all) outcome(mgale) test(cluster) nograph nolist nocheckobs replace
if (e(converged)>0) return scalar conv_stgest1_method4=1
else return scalar conv_stgest1_method4= 0
if (_rc>0) return scalar error_stgest1_method4=1
else return scalar error_stgest1_method4=0
if (e(converged)>0) {
di r(trcaus)
scalar af1 = r(trcaus)
scalar survadminc1 = admin/af1
}

restore, preserve
drop if trt==1

***Run the g-estimation on the Control group
stgest3 xoind cat1t1 cat2t1 cat3t1 cat4t1 cat1t2 cat2t2 cat3t2 cat4t2 cat1t3 cat2t3 cat3t3 cat4t3 adlag, visit(visit) lasttime(admin) range (-5 5) model(all) outcome(mgale) test(cluster) nograph nolist nocheckobs replace
if (e(converged)>0) return scalar conv_stgest0_method4=1
else return scalar conv_stgest0_method4= 0
if (_rc>0) return scalar error_stgest0_method4=1
else return scalar error_stgest0_method4=0
if (e(converged)>0) {
di r(trcaus)
scalar af0 = r(trcaus)
scalar survadminc0 = admin/af0
}

** Adjust stime and event usnig the AF generated from the g-estimation**
restore
sort id
collapse (max) trt age stime event admin xotime xoind, by(id)

*** Control group
gen cfact = (xotime + ((stime-xotime)/(af0))) if (trt==0 & xoind==1)
replace cfact = stime if (trt==0 & xoind==0)
gen dcfact = event if trt==0
replace dcfact=0 if (cfact>admin & trt==0)
replace cfact = admin if (cfact>admin & trt==0)

*** Exp group
replace cfact = (xotime + ((stime-xotime)/(af1))) if (trt==1 & xoind==1)
replace cfact = stime if (trt==1 & xoind==0)
replace dcfact = event if trt==1
replace dcfact=0 if (cfact>admin & trt==1)
replace cfact = admin if (cfact>admin & trt==1)

***Stset the data specifying the exit time to 1 year (follow-up time)
stset cfact, failure(dcfact) id(id)

capture stcox trt age
```

```
		if (e(converged)>0) return scalar conv_hr_method4=1
		else return scalar conv_hr_method4= 0
		if (_rc>0) return scalar error_hr_method4=1
		else return scalar error_hr_method4=0

		if (e(converged)>0) {
			return scalar hr_method4=exp(_b[trt])
					return scalar hr_se_method4=exp(_b[trt])*_se[trt]
		}
		else {
			return scalar hr_method4=.
					return scalar hr_se_method4=.
		}

		capture stpm2 trt age, scale(h) df(2) lininit nolog eform vce(robust)
		if (e(converged)>0) return scalar conv_rmst_method4=1
		else return scalar conv_rmst_method4= 0
		if (_rc>0) return scalar error_rmst_method4=1
		else return scalar error_rmst_method4=0

		if (e(converged)>0) {
				gen tt1 = 1 in 1
				standsurv, at1(trt 0) at2(trt 1) ci se timevar(tt1) contrast(difference) rmst atvars(rmst_trt0 
rmst_trt1) contrastvar(rmst_diff)
				summ rmst_diff, meanonly
				return scalar rmstdiff_method4= r(mean)
				summ rmst_diff_se, meanonly
				return scalar rmstdiff_se_method4= r(mean)
				summ rmst_diff_lci, meanonly
				return scalar rmstdiff_lci_method4= r(mean)
				summ rmst_diff_uci, meanonly
				return scalar rmstdiff_uci_method4= r(mean)
				summ rmst_trt0, meanonly
				return scalar rmst0_method4= r(mean)
				summ rmst_trt0_se, meanonly
				return scalar rmst_trt0_se_method4=r(mean)
				summ rmst_trt0_lci, meanonly
				return scalar rmst0_lci_method4= r(mean)
				summ rmst_trt0_uci, meanonly
				return scalar rmst0_uci_method4= r(mean)
				summ rmst_trt1, meanonly
				return scalar rmst1_method4= r(mean)
				summ rmst_trt0_se, meanonly
				return scalar rmst_trt1_se_method4=r(mean)
				summ rmst_trt1_lci, meanonly
				return scalar rmst1_lci_method4= r(mean)
				summ rmst_trt1_uci, meanonly
				return scalar rmst1_uci_method4= r(mean)
		}
		else {
				return scalar rmstdiff_method4=.
				return scalar rmstdiff_se_method4=.
				return scalar rmstdiff_lci_method4=.
				return scalar rmstdiff_uci_method4=.
				return scalar rmst0_method4=.
				return scalar rmst_trt0_se_method4=.
				return scalar rmst0_lci_method4=.
				return scalar rmst0_uci_method4=.
				return scalar rmst1_method4=.
				return scalar rmst_trt1_se_method4=.
				return scalar rmst1_lci_method4=.
				return scalar rmst1_uci_method4=.
		}
}
end

set rng mt64s		// set the  stream 64-bit Mersenne Twister
set rngstream 11	//set the stream of rng
simulate hr_method1=r(hr_method1)  ///
hr_se_method1=r(hr_se_method1) ///
conv_hr_method1=r(conv_hr_method1) ///
```

```
error_hr_method1=r(error_hr_method1) ///
rmstdiff_method1= r(rmstdiff_method1) ///
rmstdiff_se_method1=r(rmstdiff_se_method1) ///
rmstdiff_lci_method1= r(rmstdiff_lci_method1) ///
rmstdiff_uci_method1= r(rmstdiff_uci_method1) ///
rmst0_method1= r(rmst0_method1) ///
rmst_trt0_se_method1= r(rmst_trt0_se_method1) ///
rmst0_lci_method1= r(rmst0_lci_method1) ///
rmst0_uci_method1= r(rmst0_uci_method1) ///
rmst1_method1= r(rmst1_method1) ///
rmst_trt1_se_method1= r(rmst_trt1_se_method1) ///
rmst1_lci_method1= r(rmst1_lci_method1) ///
rmst1_uci_method1= r(rmst1_uci_method1) ///
conv_rmst_method1=r(conv_rmst_method1) ///
error_rmst_method1=r(error_rmst_method1) ///
hr_method2=r(hr_method2)  ///
hr_se_method2=r(hr_se_method2) ///
conv_hr_method2=r(conv_hr_method2) ///
error_hr_method2=r(error_hr_method2) ///
rmstdiff_method2= r(rmstdiff_method2) ///
rmstdiff_se_method2=r(rmstdiff_se_method2) ///
rmstdiff_lci_method2= r(rmstdiff_lci_method2) ///
rmstdiff_uci_method2= r(rmstdiff_uci_method2) ///
rmst0_method2= r(rmst0_method2) ///
rmst_trt0_se_method2= r(rmst_trt0_se_method2) ///
rmst0_lci_method2= r(rmst0_lci_method2) ///
rmst0_uci_method2= r(rmst0_uci_method2) ///
rmst1_method2= r(rmst1_method2) ///
rmst_trt1_se_method2= r(rmst_trt1_se_method2) ///
rmst1_lci_method2= r(rmst1_lci_method2) ///
rmst1_uci_method2= r(rmst1_uci_method2) ///
conv_rmst_method2=r(conv_rmst_method2) ///
error_rmst_method2=r(error_rmst_method2) ///
conv_mna1_method3=r(conv_mna1_method3) ///
error_mna1_method3=r(error_mna1_method3) ///
conv_mna2_method3=r(conv_mna2_method3) ///
error_mna2_method3=r(error_mna2_method3) ///
conv_mna3_method3=r(conv_mna3_method3) ///
error_mna3_method3=r(error_mna3_method3) ///
conv_mna4_method3=r(conv_mna4_method3) ///
error_mna4_method3=r(error_mna4_method3) ///
hr_method3=r(hr_method3)  ///
hr_se_method3=r(hr_se_method3) ///
conv_hr_method3=r(conv_hr_method3) ///
error_hr_method3=r(error_hr_method3) ///
rmstdiff_method3= r(rmstdiff_method3) ///
rmstdiff_se_method3=r(rmstdiff_se_method3) ///
rmstdiff_lci_method3= r(rmstdiff_lci_method3) ///
rmstdiff_uci_method3= r(rmstdiff_uci_method3) ///
rmst0_method3= r(rmst0_method3) ///
rmst_trt0_se_method3= r(rmst_trt0_se_method3) ///
rmst0_lci_method3= r(rmst0_lci_method3) ///
rmst0_uci_method3= r(rmst0_uci_method3) ///
rmst1_method3= r(rmst1_method3) ///
rmst_trt1_se_method3= r(rmst_trt1_se_method3) ///
rmst1_lci_method3= r(rmst1_lci_method3) ///
rmst1_uci_method3= r(rmst1_uci_method3) ///
conv_rmst_method3=r(conv_rmst_method3) ///
error_rmst_method3=r(error_rmst_method3) ///
conv_stgest1_method4=r(conv_stgest1_method4) ///
error_stgest1_method4=r(error_stgest1_method4) ///
conv_stgest0_method4=r(conv_stgest0_method4) ///
error_stgest0_method4=r(error_stgest0_method4) ///
hr_method4=r(hr_method4)  ///
hr_se_method4=r(hr_se_method4) ///
conv_hr_method4=r(conv_hr_method4) ///
error_hr_method4=r(error_hr_method4) ///
rmstdiff_method4= r(rmstdiff_method4) ///
rmstdiff_se_method4=r(rmstdiff_se_method4) ///
rmstdiff_lci_method4= r(rmstdiff_lci_method4) ///
rmstdiff_uci_method4= r(rmstdiff_uci_method4) ///
```

```
rmst0_method4= r(rmst0_method4) ///
rmst_trt0_se_method4= r(rmst_trt0_se_method4) ///
rmst0_lci_method4= r(rmst0_lci_method4) ///
rmst0_uci_method4= r(rmst0_uci_method4) ///
rmst1_method4= r(rmst1_method4) ///
rmst_trt1_se_method4= r(rmst_trt1_se_method4) ///
rmst1_lci_method4= r(rmst1_lci_method4) ///
rmst1_uci_method4= r(rmst1_uci_method4) ///
conv_rmst_method4=r(conv_rmst_method4) ///
error_rmst_method4=r(error_rmst_method4), ///
reps(1900) seed(13183) saving(estimatesc1, replace): mysimc1

use estimatesc1, clear
gen idrep= _n   // generate idrep number
order idrep, first

***reshape estimates data to long format
reshape long hr_method hr_se_method conv_hr_method error_hr_method ///
rmstdiff_method rmstdiff_se_method rmstdiff_lci_method rmstdiff_uci_method rmst0_method rmst_trt0_se_method ///
rmst0_lci_method rmst0_uci_method rmst1_method rmst_trt1_se_method rmst1_lci_method rmst1_uci_method ///
conv_mna1_method error_mna1_method conv_mna2_method error_mna2_method conv_mna3_method ///
error_mna3_method conv_mna4_method error_mna4_method conv_stgest1_method error_stgest1_method ///
conv_stgest0_method error_stgest0_method conv_rmst_method error_rmst_method, i(idrep) j(method)

***rename variable names to sensible names
rename hr_method hr
rename hr_se_method hr_se
rename conv_hr_method conv_hr
rename error_hr_method error_hr
rename rmstdiff_method rmstdiff
rename rmstdiff_se_method rmstdiff_se
rename rmstdiff_lci_method rmstdiff_lci
rename rmstdiff_uci_method rmstdiff_uci
rename rmst0_method rmst0
rename rmst_trt0_se_method rmst0_se
rename rmst0_lci_method rmst0_lci
rename rmst0_uci_method rmst0_uci
rename rmst1_method rmst1
rename rmst_trt1_se_method rmst1_se
rename rmst1_lci_method rmst1_lci
rename rmst1_uci_method rmst1_uci
rename conv_mna1_method conv_mna1
rename error_mna1_method error_mna1
rename conv_mna2_method conv_mna2
rename error_mna2_method error_mna2
rename conv_mna3_method conv_mna3
rename error_mna3_method error_mna3
rename conv_mna4_method conv_mna4
rename error_mna4_method error_mna4
rename conv_stgest1_method conv_stgest1
rename error_stgest1_method error_stgest1
rename conv_stgest0_method conv_stgest0
rename error_stgest0_method error_stgest0
rename conv_rmst_method conv_rmst
rename error_rmst_method error_rmst

*** Order vars
order rmstdiff_se, after(rmstdiff)
order hr, after(rmst1_uci)
order hr_se, after(hr)
order conv_stgest1 error_stgest1 conv_stgest0 error_stgest0, after(error_mna4)
order conv_hr, after(conv_rmst)
order error_hr, after(conv_hr)

*** Label variables and values
label variable idrep "Rep num"
label variable method "Method"
label variable hr "Hazard Ratio"
label variable hr_se "Standard Error of Hazard Ratio"
label variable rmstdiff "Difference in RMST"
label variable rmstdiff_se "Standard Error of the Difference in RMST"
```

```
label variable rmstdiff_lci "RMST 95% CI: Upper bound"
label variable rmstdiff_uci "RMST 95% CI: Lower bound"
label variable rmst0 "Restricted Mean Survival Time 'Control Group'"
label variable rmst0_se "Standard Error of RMST 'Control Group'"
label variable rmst0_lci "RMST 95% CI: Lower bound 'Control Group'"
label variable rmst0_uci "RMST 95% CI: Upper bound 'Control Group'"
label variable rmst1 "Restricted Mean Survival Time 'Exp Group'"
label variable rmst1_se "Standard Error of RMST 'Exp Group'"
label variable rmst1_lci "RMST 95% CI: Lower bound 'Exp Group'"
label variable rmst1_uci "RMST 95% CI: Upper bound 'Exp Group'"
label variable conv_hr "HR model converged"
label variable error_hr "Error - HR model"
label variable conv_rmst "RMST model converged"
label variable error_rmst "Error - RMST model"
label variable conv_mna1 "Non-adherence model(1) converged"
label variable error_mna1 "Error - Non-adherence model(1)"
label variable conv_mna2 "Non-adherence model(2) converged"
label variable error_mna2 "Error - Non-adherence model(2)"
label variable conv_mna3 "Non-adherence model(3) converged"
label variable error_mna3 "Error - Non-adherence model(3)"
label variable conv_mna4 "Non-adherence model(4) converged"
label variable error_mna4 "Error - Non-adherence model(4)"
label variable conv_stgest1 "G-estimation converged 'Exp Group'"
label variable error_stgest1 "Error - G-estimation 'Exp Group'"
label variable conv_stgest0 "G-estimation converged 'Control Group'"
label variable error_stgest0 "Error - G-estimation 'Control Group'"
label define nylab 0 "No" 1 "Yes"
label values conv_hr conv_rmst error_hr error_rmst nylab
label values conv_mna1 error_mna1 conv_mna2 error_mna2 conv_mna3 error_mna3 nylab
label values conv_mna4 error_mna4 conv_stgest1 error_stgest1 conv_stgest0 error_stgest0 nylab
label define methodlab 1 "ITT" 2 "PP" 3 "IPCW"  4 "SNFTM"
label values method methodlab

***Save labelled estimates dataset
save adjusted1, replace
```

Supplementary Appendix D: Performance of methods across scenarios

Table 5: Performance of methods across implementation non-adherence scenarios

Scenario No.	Successful nsim	Truth (Diff. in RMSTs)	Method	Mean estimate (Diff. in RMSTs)	SE of mean	95% Confidence interval		Bias	Bias (%)	MSE (%)	Model SE (%)	Empirical SE (%)	Coverage (%)	Successful estimation (%)
						Lower	Upper							
1	1872	0.108	ITT	0.135	0.032	0.072	0.197	0.026	24.26	1.54	29.37	28.86	86.49	100.00
			PP	0.131	0.037	0.059	0.202	0.022	20.61	1.68	33.74	33.53	90.33	100.00
			IPCW	0.128	0.037	0.056	0.201	0.020	18.24	1.62	34.05	34.14	91.08	100.00
			SNFTM	0.129	0.031	0.067	0.190	0.020	18.52	1.84	28.81	36.85	83.49	100.00
			min MCSE	-	-	-	-	-	0.07	0.00	0.00	0.05	0.66	-
			max MCSE	-	-	-	-	-	0.09	0.01	0.00	0.07	0.86	-
2	1874	0.063	ITT	0.105	0.030	0.046	0.164	0.042	66.06	4.18	47.48	47.26	71.08	100.00
			PP	0.083	0.033	0.019	0.146	0.019	30.43	2.29	51.47	51.82	90.72	100.00
			IPCW	0.084	0.033	0.018	0.149	0.020	31.94	2.43	52.57	53.09	90.72	100.00
			SNFTM	0.082	0.028	0.028	0.136	0.019	30.07	2.40	43.46	53.71	82.55	100.00
			min MCSE	-	-	-	-	-	0.07	0.00	0.00	0.05	0.67	-
			max MCSE	-	-	-	-	-	0.08	0.01	0.00	0.06	1.05	-
3	1863	0.108	ITT	0.136	0.031	0.075	0.196	0.027	25.18	1.56	28.49	28.39	85.72	100.00
			PP	0.124	0.034	0.056	0.192	0.016	14.34	1.27	31.80	31.13	93.34	100.00
			IPCW	0.123	0.035	0.054	0.192	0.015	13.66	1.34	32.50	32.38	92.22	100.00
			SNFTM	0.121	0.031	0.061	0.181	0.013	11.60	1.41	28.22	34.20	88.35	100.00
			min MCSE	-	-	-	-	-	0.07	0.00	0.00	0.05	0.66	-
			max MCSE	-	-	-	-	-	0.09	0.01	0.00	0.07	0.86	-
4	1855	0.063	ITT	0.091	0.027	0.038	0.145	0.028	44.12	2.34	42.90	41.74	83.29	100.00
			PP	0.074	0.029	0.016	0.132	0.010	16.54	1.51	46.53	45.93	93.96	100.00
			IPCW	0.074	0.030	0.015	0.134	0.011	17.49	1.62	48.06	47.48	94.07	100.00
			SNFTM	0.072	0.026	0.021	0.122	0.008	13.34	1.49	40.82	46.58	90.84	100.00
			min MCSE	-	-	-	-	-	0.06	0.00	0.00	0.04	0.55	-
			max MCSE	-	-	-	-	-	0.07	0.00	0.00	0.05	0.87	-
5	1859	0.108	ITT	0.131	0.031	0.070	0.192	0.022	20.51	1.33	28.70	28.34	88.81	100.00
			PP	0.123	0.036	0.052	0.194	0.015	13.57	1.38	33.52	33.05	92.79	100.00
			IPCW	0.121	0.037	0.049	0.193	0.012	11.42	1.36	33.92	33.57	93.01	100.00

Scenario No.	Successful nsim	Truth (Diff. in RMSTs)	Method	Mean estimate (Diff. in RMSTs)	SE of mean	95% Confidence interval		Bias	Bias (%)	MSE (%)	Model SE (%)	Empirical SE (%)	Coverage (%)	Successful estimation (%)
						Lower	Upper							
			SNFTM	0.120	0.031	0.060	0.180	0.012	10.74	1.65	28.33	37.52	84.51	100.00
			min MCSE	-	-	-	-	-	0.07	0.00	0.00	0.05	0.59	-
			max MCSE	-	-	-	-	-	0.09	0.01	0.00	0.07	0.84	-
6	1845	0.063	ITT	0.091	0.028	0.036	0.146	0.028	43.56	2.41	44.09	43.63	83.31	100.00
			PP	0.073	0.031	0.012	0.135	0.010	15.79	1.72	49.49	49.66	93.82	100.00
			IPCW	0.075	0.032	0.011	0.138	0.011	17.74	1.86	51.05	51.20	93.88	100.00
			SNFTM	0.072	0.026	0.021	0.124	0.009	14.41	1.80	41.26	51.38	87.59	100.00
			min MCSE	-	-	-	-	-	0.06	0.00	0.00	0.05	0.56	-
			max MCSE	-	-	-	-	-	0.08	0.00	0.00	0.05	0.87	-
7	1840	0.169	ITT	0.209	0.031	0.148	0.270	0.040	23.94	1.52	18.39	18.08	75.22	100.00
			PP	0.198	0.034	0.131	0.264	0.029	17.29	1.18	20.13	19.96	86.63	100.00
			IPCW	0.196	0.035	0.128	0.264	0.027	16.07	1.16	20.57	20.75	87.61	100.00
			SNFTM	0.196	0.030	0.136	0.255	0.027	16.02	1.17	17.93	20.94	82.17	100.00
			min MCSE	-	-	-	-	-	0.07	0.01	0.00	0.05	0.77	-
			max MCSE	-	-	-	-	-	0.08	0.01	0.00	0.06	1.01	-
8	1841	0.095	ITT	0.148	0.028	0.094	0.203	0.054	56.71	3.88	29.49	29.63	51.33	100.00
			PP	0.120	0.029	0.062	0.177	0.025	26.28	1.54	31.01	30.57	87.34	100.00
			IPCW	0.120	0.030	0.061	0.179	0.025	26.77	1.62	31.79	31.53	87.51	100.00
			SNFTM	0.119	0.026	0.068	0.170	0.024	25.64	1.55	27.40	31.31	81.31	100.00
			min MCSE	-	-	-	-	-	0.07	0.00	0.00	0.05	0.77	-
			max MCSE	-	-	-	-	-	0.07	0.01	0.00	0.05	1.16	-
9	1845	0.169	ITT	0.202	0.031	0.140	0.263	0.033	19.65	1.25	18.65	18.82	81.52	100.00
			PP	0.198	0.036	0.128	0.268	0.029	17.34	1.25	21.27	20.94	87.37	100.00
			IPCW	0.194	0.036	0.123	0.265	0.026	15.26	1.17	21.46	21.46	89.05	100.00
			SNFTM	0.195	0.030	0.136	0.255	0.027	15.79	1.34	18.07	23.41	80.05	100.00
			min MCSE	-	-	-	-	-	0.07	0.01	0.00	0.05	0.73	-
			max MCSE	-	-	-	-	-	0.09	0.01	0.00	0.06	0.93	-
10	1822	0.095	ITT	0.150	0.029	0.094	0.207	0.056	58.88	4.14	30.59	30.15	51.54	100.00
			PP	0.120	0.031	0.059	0.181	0.025	26.69	1.70	33.09	32.97	86.66	100.00
			IPCW	0.121	0.032	0.058	0.184	0.026	27.61	1.82	33.80	34.08	85.84	100.00

Scenario No.	Successful nsim	Truth (Diff. in RMSTs)	Method	Mean estimate (Diff. in RMSTs)	SE of mean	95% Confidence interval		Bias	Bias (%)	MSE (%)	Model SE (%)	Empirical SE (%)	Coverage (%)	Successful estimation (%)
						Lower	Upper							
			SNFTM	0.120	0.026	0.068	0.171	0.025	26.25	1.90	27.85	36.32	76.89	100.00
			min MCSE	-	-	-	-	-	0.07	0.00	0.00	0.05	0.80	-
			max MCSE	-	-	-	-	-	0.08	0.01	0.00	0.06	1.17	-
11	1826	0.155	ITT	0.195	0.031	0.135	0.256	0.040	26.04	1.63	20.02	19.27	74.92	100.00
			PP	0.185	0.034	0.119	0.252	0.030	19.46	1.28	21.93	21.08	86.04	100.00
			IPCW	0.183	0.035	0.115	0.251	0.028	18.08	1.28	22.39	22.38	87.51	100.00
			SNFTM	0.182	0.030	0.122	0.241	0.027	17.33	1.40	19.56	24.62	79.30	100.00
			min MCSE	-	-	-	-	-	0.07	0.01	0.00	0.05	0.77	-
			max MCSE	-	-	-	-	-	0.09	0.01	0.00	0.06	1.01	-
12	1829	0.087	ITT	0.141	0.028	0.086	0.196	0.054	61.47	4.21	32.11	32.20	51.67	100.00
			PP	0.113	0.029	0.055	0.171	0.026	29.38	1.79	33.75	34.42	84.96	100.00
			IPCW	0.114	0.030	0.054	0.173	0.026	29.87	1.89	34.62	35.63	84.64	100.00
			SNFTM	0.112	0.026	0.060	0.163	0.024	27.57	1.82	29.83	36.30	79.33	100.00
			min MCSE	-	-	-	-	-	0.07	0.00	0.00	0.05	0.84	-
			max MCSE	-	-	-	-	-	0.07	0.01	0.00	0.05	1.17	-
13	1855	0.155	ITT	0.188	0.031	0.126	0.250	0.033	21.30	1.31	20.26	19.80	82.32	100.00
			PP	0.185	0.036	0.115	0.256	0.030	19.64	1.38	23.17	22.43	86.85	100.00
			IPCW	0.182	0.036	0.111	0.253	0.027	17.48	1.31	23.36	23.27	88.73	100.00
			SNFTM	0.182	0.031	0.122	0.242	0.027	17.49	1.47	19.69	25.39	78.81	100.00
			min MCSE	-	-	-	-	-	0.07	0.01	0.00	0.05	0.73	-
			max MCSE	-	-	-	-	-	0.09	0.01	0.00	0.06	0.95	-
14	1825	0.087	ITT	0.143	0.029	0.086	0.200	0.055	63.10	4.47	33.29	33.62	51.40	100.00
			PP	0.114	0.031	0.053	0.176	0.027	30.80	2.04	36.01	37.17	84.60	100.00
			IPCW	0.115	0.032	0.052	0.178	0.028	31.90	2.17	36.83	38.29	85.04	100.00
			SNFTM	0.113	0.026	0.061	0.165	0.026	29.59	2.22	30.28	40.84	74.30	100.00
			min MCSE	-	-	-	-	-	0.07	0.01	0.00	0.05	0.83	-
			max MCSE	-	-	-	-	-	0.08	0.01	0.00	0.06	1.17	-
15	1801	0.093	ITT	0.121	0.031	0.061	0.182	0.028	30.44	1.94	33.18	34.06	83.73	100.00
			PP	0.110	0.034	0.043	0.178	0.017	18.71	1.70	37.07	38.42	90.62	100.00
			IPCW	0.110	0.035	0.041	0.180	0.017	18.72	1.78	37.90	39.51	90.34	100.00

Scenario No.	Successful nsim	Truth (Diff. in RMSTs)	Method	Mean estimate (Diff. in RMSTs)	SE of mean	95% Confidence interval		Bias	Bias (%)	MSE (%)	Model SE (%)	Empirical SE (%)	Coverage (%)	Successful estimation (%)
						Lower	Upper							
			SNFTM	0.107	0.031	0.047	0.167	0.014	15.05	2.02	32.93	44.14	83.95	100.00
			min MCSE	-	-	-	-	-	0.07	0.01	0.00	0.05	0.69	-
			max MCSE	-	-	-	-	-	0.10	0.01	0.00	0.07	0.87	-
16	1811	0.054	ITT	0.081	0.027	0.028	0.135	0.027	49.10	2.77	50.23	51.79	82.11	100.00
			PP	0.065	0.030	0.007	0.123	0.011	19.81	1.90	54.50	55.64	92.88	100.00
			IPCW	0.065	0.031	0.006	0.125	0.011	20.33	2.03	56.17	57.60	93.21	100.00
			SNFTM	0.064	0.026	0.013	0.114	0.009	16.77	2.03	47.74	58.76	87.24	100.00
			min MCSE	-	-	-	-	-	0.07	0.00	0.00	0.05	0.59	-
			max MCSE	-	-	-	-	-	0.08	0.00	0.00	0.05	0.90	-
17	1854	0.093	ITT	0.116	0.031	0.055	0.177	0.023	24.78	1.69	33.42	34.68	86.79	100.00
			PP	0.110	0.036	0.039	0.181	0.017	17.94	1.83	39.05	40.50	91.48	100.00
			IPCW	0.108	0.037	0.036	0.180	0.015	15.87	1.83	39.54	41.36	91.91	100.00
			SNFTM	0.105	0.031	0.045	0.166	0.012	13.07	2.33	33.08	48.36	81.45	100.00
			min MCSE	-	-	-	-	-	0.07	0.00	0.00	0.05	0.63	-
			max MCSE	-	-	-	-	-	0.10	0.01	0.00	0.07	0.90	-
18	1803	0.054	ITT	0.081	0.028	0.026	0.136	0.027	49.22	2.80	51.57	52.30	84.03	100.00
			PP	0.065	0.031	0.003	0.127	0.011	19.71	2.17	57.88	59.98	92.62	100.00
			IPCW	0.066	0.032	0.003	0.130	0.012	21.80	2.34	59.58	61.94	92.68	100.00
			SNFTM	0.064	0.026	0.013	0.116	0.010	18.37	2.58	48.15	66.38	83.47	100.00
			min MCSE	-	-	-	-	-	0.07	0.00	0.00	0.05	0.61	-
			max MCSE	-	-	-	-	-	0.09	0.00	0.00	0.06	0.87	-
19	1625	0.108	ITT	0.146	0.059	0.031	0.261	0.038	34.69	4.39	54.26	53.37	90.65	100.00
			PP	0.134	0.065	0.007	0.261	0.026	23.72	4.38	59.71	58.99	92.98	100.00
			IPCW	0.133	0.081	-0.025	0.291	0.025	22.91	7.01	75.13	77.12	92.43	100.00
			SNFTM	0.133	0.058	0.020	0.246	0.024	22.43	4.74	53.33	62.24	88.55	100.00
			min MCSE	-	-	-	-	-	0.14	0.02	0.01	0.10	0.63	-
			max MCSE	-	-	-	-	-	0.21	0.03	0.03	0.15	0.79	-
20	1773	0.064	ITT	0.111	0.057	0.000	0.222	0.047	74.61	8.41	89.19	87.62	87.65	100.00
			PP	0.086	0.060	-0.031	0.203	0.022	35.25	6.19	93.96	92.24	93.74	100.00
			IPCW	0.087	0.073	-0.056	0.230	0.024	37.26	9.42	116.33	115.96	93.17	100.00

Scenario No.	Successful nsim	Truth (Diff. in RMSTs)	Method	Mean estimate (Diff. in RMSTs)	SE of mean	95% Confidence interval Lower	Upper	Bias	Bias (%)	MSE (%)	Model SE (%)	Empirical SE (%)	Coverage (%)	Successful estimation (%)
			SNFTM	0.088	0.052	-0.015	0.191	0.024	38.13	6.70	82.77	95.38	87.70	100.00
			min MCSE	-	-	-	-	-	0.13	0.01	0.01	0.09	0.58	-
			max MCSE	-	-	-	-	-	0.18	0.02	0.03	0.12	0.78	-
21	1671	0.108	ITT	0.134	0.062	0.013	0.255	0.026	23.57	3.95	56.97	55.63	92.22	100.00
			PP	0.130	0.071	-0.009	0.269	0.022	20.29	4.79	65.54	63.33	94.61	100.00
			IPCW	0.126	0.084	-0.038	0.290	0.018	16.25	6.91	77.87	78.23	94.14	100.00
			SNFTM	0.128	0.060	0.010	0.246	0.020	18.22	5.33	55.81	67.79	87.49	100.00
			min MCSE	-	-	-	-	-	0.15	0.01	0.01	0.10	0.55	-
			max MCSE	-	-	-	-	-	0.21	0.02	0.02	0.15	0.81	-
22	1776	0.064	ITT	0.107	0.059	-0.008	0.222	0.044	68.52	8.31	92.41	91.58	88.40	100.00
			PP	0.085	0.064	-0.040	0.209	0.021	33.54	6.88	100.40	98.53	92.91	100.00
			IPCW	0.086	0.074	-0.059	0.232	0.023	36.06	9.60	118.64	117.56	93.35	100.00
			SNFTM	0.085	0.053	-0.019	0.190	0.022	34.36	7.46	84.15	102.80	85.92	100.00
			min MCSE	-	-	-	-	-	0.14	0.01	0.01	0.10	0.59	-
			max MCSE	-	-	-	-	-	0.18	0.02	0.04	0.13	0.83	-
23	1697	0.108	ITT	0.132	0.060	0.015	0.249	0.024	22.26	3.62	55.15	53.41	93.93	100.00
			PP	0.120	0.067	-0.011	0.251	0.012	11.10	4.09	61.80	60.48	95.05	100.00
			IPCW	0.117	0.083	-0.045	0.279	0.009	8.11	6.91	76.83	79.47	92.52	100.00
			SNFTM	0.117	0.059	0.001	0.233	0.009	7.94	4.32	54.57	62.65	90.93	100.00
			min MCSE	-	-	-	-	-	0.14	0.01	0.01	0.10	0.53	-
			max MCSE	-	-	-	-	-	0.21	0.03	0.03	0.15	0.70	-
24	1785	0.064	ITT	0.092	0.053	-0.012	0.196	0.029	45.37	5.75	83.84	83.66	91.32	100.00
			PP	0.076	0.058	-0.037	0.188	0.012	18.92	5.50	90.89	91.14	94.17	100.00
			IPCW	0.076	0.071	-0.063	0.214	0.012	19.24	8.48	112.64	113.96	93.61	100.00
			SNFTM	0.076	0.050	-0.023	0.174	0.012	19.15	5.79	79.29	93.52	89.41	100.00
			min MCSE	-	-	-	-	-	0.13	0.01	0.01	0.09	0.55	-
			max MCSE	-	-	-	-	-	0.17	0.02	0.03	0.12	0.73	-
25	1739	0.108	ITT	0.130	0.060	0.012	0.248	0.022	20.22	3.64	55.69	54.32	93.27	100.00
			PP	0.121	0.070	-0.017	0.259	0.013	11.86	4.60	65.17	64.10	94.59	100.00
			IPCW	0.116	0.082	-0.045	0.277	0.008	7.11	6.62	76.44	77.87	94.08	100.00

Scenario No.	Successful nsim	Truth (Diff. in RMSTs)	Method	Mean estimate (Diff. in RMSTs)	SE of mean	95% Confidence interval Lower	Upper	Bias	Bias (%)	MSE (%)	Model SE (%)	Empirical SE (%)	Coverage (%)	Successful estimation (%)
			SNFTM	0.120	0.059	0.004	0.236	0.012	10.82	5.43	54.88	69.99	87.18	100.00
			min MCSE	-	-	-	-	-	0.14	0.01	0.01	0.10	0.54	-
			max MCSE	-	-	-	-	-	0.20	0.02	0.02	0.14	0.80	-
26	1806	0.064	ITT	0.092	0.055	-0.015	0.199	0.029	45.19	5.98	85.99	85.84	91.36	100.00
			PP	0.075	0.061	-0.045	0.195	0.011	17.80	5.96	96.52	95.20	94.13	100.00
			IPCW	0.077	0.071	-0.063	0.217	0.013	21.00	8.30	113.50	112.36	94.46	100.00
			SNFTM	0.076	0.051	-0.024	0.175	0.012	19.16	6.43	80.08	98.81	87.38	100.00
			min MCSE	-	-	-	-	-	0.13	0.01	0.01	0.09	0.54	-
			max MCSE	-	-	-	-	-	0.17	0.02	0.03	0.12	0.78	-
27	1622	0.169	ITT	0.210	0.060	0.093	0.328	0.042	24.81	3.18	35.56	35.64	88.10	100.00
			PP	0.199	0.066	0.070	0.327	0.030	17.77	3.08	39.04	38.92	91.68	100.00
			IPCW	0.194	0.082	0.033	0.355	0.025	15.10	4.70	49.14	50.61	91.55	100.00
			SNFTM	0.196	0.058	0.082	0.310	0.028	16.37	3.16	34.61	40.07	87.85	100.00
			min MCSE	-	-	-	-	-	0.15	0.02	0.01	0.11	0.69	-
			max MCSE	-	-	-	-	-	0.21	0.03	0.03	0.15	0.81	-
28	1801	0.095	ITT	0.149	0.054	0.043	0.256	0.054	56.94	6.04	57.35	55.72	83.73	100.00
			PP	0.119	0.057	0.008	0.231	0.024	25.40	3.93	60.19	59.07	92.73	100.00
			IPCW	0.120	0.070	-0.017	0.257	0.025	26.05	5.66	74.42	72.58	94.16	100.00
			SNFTM	0.120	0.050	0.022	0.219	0.025	26.53	4.19	53.03	60.86	88.01	100.00
			min MCSE	-	-	-	-	-	0.12	0.01	0.01	0.09	0.55	-
			max MCSE	-	-	-	-	-	0.16	0.02	0.03	0.12	0.87	-
29	1727	0.169	ITT	0.203	0.061	0.084	0.322	0.034	20.39	2.96	36.05	36.62	90.21	100.00
			PP	0.198	0.069	0.063	0.334	0.030	17.71	3.38	41.19	41.11	91.60	100.00
			IPCW	0.191	0.082	0.030	0.351	0.022	13.02	4.57	48.76	50.42	91.84	100.00
			SNFTM	0.195	0.059	0.080	0.310	0.027	15.78	3.54	34.85	43.01	84.83	100.00
			min MCSE	-	-	-	-	-	0.15	0.02	0.01	0.11	0.66	-
			max MCSE	-	-	-	-	-	0.20	0.03	0.02	0.14	0.86	-
30	1819	0.095	ITT	0.151	0.056	0.041	0.262	0.056	58.73	6.63	59.34	59.32	83.34	100.00
			PP	0.120	0.061	0.001	0.240	0.025	26.55	4.60	64.22	64.30	92.58	100.00
			IPCW	0.122	0.071	-0.017	0.261	0.027	28.42	6.35	75.06	76.58	92.35	100.00

Scenario No.	Successful nsim	Truth (Diff. in RMSTs)	Method	Mean estimate (Diff. in RMSTs)	SE of mean	95% Confidence interval Lower	Upper	Bias	Bias (%)	MSE (%)	Model SE (%)	Empirical SE (%)	Coverage (%)	Successful estimation (%)
			SNFTM	0.122	0.051	0.022	0.222	0.027	28.18	4.95	53.90	66.43	86.64	100.00
			min MCSE	-	-	-	-	-	0.13	0.01	0.01	0.09	0.61	-
			max MCSE	-	-	-	-	-	0.17	0.02	0.02	0.12	0.87	-
31	1608	0.155	ITT	0.193	0.060	0.075	0.311	0.038	24.30	3.32	38.79	39.38	89.49	100.00
			PP	0.183	0.066	0.054	0.313	0.028	18.25	3.41	42.55	43.21	91.79	100.00
			IPCW	0.178	0.083	0.015	0.340	0.023	14.65	5.44	53.93	57.39	90.72	100.00
			SNFTM	0.179	0.059	0.065	0.294	0.024	15.71	3.57	37.86	45.33	87.56	100.00
			min MCSE	-	-	-	-	-	0.15	0.02	0.01	0.11	0.68	-
			max MCSE	-	-	-	-	-	0.22	0.03	0.03	0.16	0.82	-
32	1802	0.088	ITT	0.139	0.055	0.031	0.246	0.051	57.78	6.22	62.40	61.23	84.74	100.00
			PP	0.110	0.057	-0.003	0.222	0.022	24.90	4.21	65.53	64.65	93.17	100.00
			IPCW	0.112	0.070	-0.024	0.249	0.025	27.90	6.47	80.46	81.18	92.44	100.00
			SNFTM	0.111	0.051	0.012	0.210	0.024	26.77	4.49	57.77	66.34	88.57	100.00
			min MCSE	-	-	-	-	-	0.13	0.01	0.01	0.09	0.59	-
			max MCSE	-	-	-	-	-	0.17	0.02	0.03	0.12	0.85	-
33	1665	0.155	ITT	0.186	0.061	0.067	0.306	0.031	20.23	3.15	39.25	40.31	90.81	100.00
			PP	0.184	0.070	0.047	0.320	0.029	18.50	3.77	44.94	45.71	91.95	100.00
			IPCW	0.179	0.082	0.018	0.339	0.024	15.25	5.25	53.18	56.16	91.77	100.00
			SNFTM	0.182	0.059	0.066	0.297	0.027	17.22	3.96	38.03	47.52	85.59	100.00
			min MCSE	-	-	-	-	-	0.15	0.02	0.01	0.11	0.67	-
			max MCSE	-	-	-	-	-	0.21	0.03	0.03	0.15	0.86	-
34	1811	0.088	ITT	0.140	0.057	0.029	0.251	0.052	59.19	6.72	64.69	64.38	85.26	100.00
			PP	0.111	0.061	-0.009	0.231	0.023	26.26	4.84	69.99	69.44	93.43	100.00
			IPCW	0.114	0.071	-0.025	0.254	0.026	30.13	6.84	81.68	82.94	93.04	100.00
			SNFTM	0.111	0.051	0.011	0.212	0.023	26.74	5.23	58.58	72.37	86.20	100.00
			min MCSE	-	-	-	-	-	0.13	0.01	0.01	0.09	0.58	-
			max MCSE	-	-	-	-	-	0.17	0.02	0.02	0.12	0.83	-
35	1633	0.093	ITT	0.117	0.060	0.000	0.234	0.024	25.76	4.61	64.08	65.45	91.92	100.00
			PP	0.105	0.067	-0.026	0.235	0.012	12.61	5.26	71.66	74.12	93.63	100.00
			IPCW	0.102	0.083	-0.061	0.265	0.009	9.40	8.89	90.23	97.30	92.84	100.00

Scenario No.	Successful nsim	Truth (Diff. in RMSTs)	Method	Mean estimate (Diff. in RMSTs)	SE of mean	95% Confidence interval		Bias	Bias (%)	MSE (%)	Model SE (%)	Empirical SE (%)	Coverage (%)	Successful estimation (%)
						Lower	Upper							
			SNFTM	0.101	0.059	-0.015	0.217	0.008	8.13	5.83	63.55	78.75	87.94	100.00
			min MCSE	-	-	-	-	-	0.15	0.01	0.01	0.11	0.60	-
			max MCSE	-	-	-	-	-	0.22	0.03	0.03	0.16	0.81	-
36	1785	0.055	ITT	0.084	0.053	-0.020	0.189	0.030	54.11	6.78	97.86	97.44	91.71	100.00
			PP	0.068	0.058	-0.045	0.181	0.013	24.05	6.22	105.96	104.03	94.57	100.00
			IPCW	0.067	0.071	-0.071	0.206	0.013	23.50	9.18	131.19	127.61	95.46	100.00
			SNFTM	0.067	0.050	-0.031	0.166	0.013	23.23	6.43	92.57	106.01	90.76	100.00
			min MCSE	-	-	-	-	-	0.13	0.01	0.01	0.09	0.49	-
			max MCSE	-	-	-	-	-	0.16	0.02	0.03	0.12	0.69	-
37	1754	0.093	ITT	0.113	0.060	-0.005	0.230	0.019	20.80	4.37	64.59	65.32	93.16	100.00
			PP	0.105	0.070	-0.032	0.243	0.012	13.25	5.58	75.58	76.28	94.18	100.00
			IPCW	0.102	0.082	-0.060	0.263	0.009	9.42	8.09	89.01	92.73	93.22	100.00
			SNFTM	0.102	0.059	-0.015	0.218	0.008	9.04	5.96	63.86	79.48	86.89	100.00
			min MCSE	-	-	-	-	-	0.15	0.01	0.01	0.10	0.56	-
			max MCSE	-	-	-	-	-	0.21	0.03	0.02	0.15	0.81	-
38	1822	0.055	ITT	0.084	0.055	-0.024	0.191	0.029	53.09	7.18	100.39	101.67	91.44	100.00
			PP	0.069	0.061	-0.051	0.189	0.015	26.77	7.34	112.64	112.86	93.91	100.00
			IPCW	0.072	0.071	-0.068	0.211	0.017	31.10	10.13	132.08	132.67	93.58	100.00
			SNFTM	0.068	0.051	-0.031	0.168	0.014	24.77	7.70	93.42	116.21	87.54	100.00
			min MCSE	-	-	-	-	-	0.13	0.01	0.01	0.09	0.56	-
			max MCSE	-	-	-	-	-	0.17	0.02	0.03	0.12	0.77	-

Table 6: Performance of methods across persistence non-adherence scenarios

Scenario No.	Successful nsim	Truth (Diff. in RMSTs)	Method	Mean estimate (Diff. in RMSTs)	SE of mean	95% Confidence interval		Bias	Percent bias	MSE	Model SE	Empirical SE	Coverage (%)	Successful estimation (%)
						Lower	Upper							
39	1817	0.108	ITT	0.131	0.032	0.068	0.195	0.023	21.18	29.29	29.71	1.42	88.77	100.00
			PP	0.130	0.035	0.060	0.199	0.021	19.52	32.04	32.61	1.53	90.64	100.00
			IPCW	0.127	0.036	0.057	0.198	0.019	17.54	33.20	33.31	1.53	91.19	100.00
			SNFTM	0.128	0.031	0.067	0.190	0.020	18.34	37.35	28.94	1.88	83.32	100.00
			min MCSE	-	-	-	-	-	0.07	0.00	0.00	0.05	0.66	-
			max MCSE	-	-	-	-	-	0.10	0.01	0.00	0.07	0.87	-
40	1812	0.063	ITT	0.104	0.030	0.044	0.163	0.040	63.44	47.61	48.02	3.98	73.01	100.00
			PP	0.083	0.031	0.022	0.145	0.020	31.76	49.19	49.72	2.17	90.89	100.00
			IPCW	0.084	0.032	0.021	0.147	0.021	33.00	50.09	50.69	2.28	90.89	100.00
			SNFTM	0.083	0.028	0.029	0.138	0.020	31.64	53.45	43.74	2.44	82.73	100.00
			min MCSE	-	-	-	-	-	0.07	0.00	0.00	0.05	0.68	-
			max MCSE	-	-	-	-	-	0.08	0.01	0.00	0.06	1.04	-
41	1824	0.108	ITT	0.138	0.031	0.077	0.200	0.030	27.46	28.52	28.91	1.70	83.94	100.00
			PP	0.127	0.034	0.061	0.193	0.019	17.22	30.50	31.06	1.33	91.94	100.00
			IPCW	0.129	0.035	0.059	0.198	0.020	18.73	33.08	32.62	1.57	90.52	100.00
			SNFTM	0.125	0.031	0.064	0.185	0.016	15.21	32.62	28.46	1.40	88.43	100.00
			min MCSE	-	-	-	-	-	0.07	0.00	0.00	0.05	0.64	-
			max MCSE	-	-	-	-	-	0.08	0.01	0.00	0.06	0.86	-
42	1811	0.063	ITT	0.099	0.028	0.044	0.154	0.035	55.87	43.08	44.37	3.15	76.59	100.00
			PP	0.078	0.029	0.021	0.135	0.015	23.85	45.39	45.95	1.66	92.55	100.00
			IPCW	0.078	0.030	0.019	0.138	0.015	23.69	47.62	48.07	1.79	92.44	100.00
			SNFTM	0.078	0.026	0.026	0.129	0.014	22.57	45.95	41.72	1.66	88.85	100.00
			min MCSE	-	-	-	-	-	0.06	0.00	0.00	0.05	0.62	-
			max MCSE	-	-	-	-	-	0.07	0.01	0.00	0.05	1.00	-
43	1830	0.095	ITT	0.147	0.028	0.092	0.203	0.053	55.57	30.78	29.76	3.82	52.51	100.00
			PP	0.120	0.029	0.064	0.176	0.025	26.89	30.76	30.21	1.58	85.57	100.00
			IPCW	0.120	0.030	0.062	0.178	0.025	26.69	32.20	31.40	1.66	86.07	100.00
			SNFTM	0.120	0.026	0.069	0.171	0.025	26.68	31.81	27.53	1.63	80.11	100.00
			min MCSE	-	-	-	-	-	0.07	0.00	0.00	0.05	0.81	-

Scenario No.	Successful nsim	Truth (Diff. in RMSTs)	Method	Mean estimate (Diff. in RMSTs)	SE of mean	95% Confidence interval		Bias	Percent bias	MSE	Model SE	Empirical SE	Coverage (%)	Successful estimation (%)
						Lower	Upper							
			max MCSE	-	-	-	-	-	0.07	0.01	0.00	0.05	1.17	-
44	1809	0.169	ITT	0.195	0.032	0.133	0.257	0.027	15.78	18.72	18.84	1.01	86.07	100.00
			PP	0.196	0.035	0.128	0.263	0.027	16.04	20.32	20.54	1.13	88.23	100.00
			IPCW	0.192	0.035	0.123	0.262	0.024	14.18	21.16	20.93	1.09	89.17	100.00
			SNFTM	0.194	0.031	0.134	0.254	0.026	15.17	22.57	18.13	1.25	80.71	100.00
			min MCSE	-	-	-	-	-	0.07	0.01	0.00	0.05	0.73	-
			max MCSE	-	-	-	-	-	0.09	0.01	0.00	0.06	0.93	-
45	1813	0.095	ITT	0.148	0.029	0.090	0.206	0.053	56.25	31.27	31.00	3.92	55.16	100.00
			PP	0.121	0.030	0.062	0.181	0.027	28.17	31.70	32.03	1.70	86.38	100.00
			IPCW	0.122	0.031	0.062	0.183	0.027	28.89	32.60	32.63	1.80	85.05	100.00
			SNFTM	0.122	0.027	0.069	0.174	0.027	28.36	34.27	28.09	1.87	78.49	100.00
			min MCSE	-	-	-	-	-	0.07	0.00	0.00	0.05	0.81	-
			max MCSE	-	-	-	-	-	0.08	0.01	0.00	0.05	1.17	-
46	1804	0.155	ITT	0.192	0.031	0.130	0.253	0.037	23.60	19.26	20.17	1.44	78.88	100.00
			PP	0.183	0.033	0.119	0.248	0.028	18.34	20.44	21.38	1.17	86.92	100.00
			IPCW	0.184	0.035	0.116	0.252	0.029	18.79	22.21	22.46	1.31	86.31	100.00
			SNFTM	0.181	0.030	0.121	0.241	0.026	16.77	22.08	19.61	1.19	82.59	100.00
			min MCSE	-	-	-	-	-	0.07	0.01	0.00	0.05	0.79	-
			max MCSE	-	-	-	-	-	0.08	0.01	0.00	0.06	0.96	-
47	1827	0.087	ITT	0.140	0.028	0.084	0.195	0.052	59.65	32.72	32.39	4.05	54.30	100.00
			PP	0.114	0.029	0.057	0.170	0.026	29.81	33.03	32.89	1.73	85.77	100.00
			IPCW	0.113	0.030	0.055	0.172	0.026	29.61	34.39	34.25	1.80	86.75	100.00
			SNFTM	0.113	0.026	0.061	0.164	0.025	28.76	33.81	30.01	1.72	81.17	100.00
			min MCSE	-	-	-	-	-	0.07	0.00	0.00	0.05	0.79	-
			max MCSE	-	-	-	-	-	0.07	0.01	0.00	0.05	1.17	-
48	1818	0.155	ITT	0.183	0.032	0.120	0.245	0.028	17.85	20.14	20.50	1.12	85.31	100.00
			PP	0.183	0.035	0.115	0.251	0.028	18.05	21.62	22.39	1.23	87.02	100.00
			IPCW	0.180	0.035	0.111	0.249	0.025	16.10	22.64	22.83	1.20	89.05	100.00
			SNFTM	0.181	0.031	0.121	0.241	0.026	16.82	25.03	19.79	1.41	80.42	100.00
			min MCSE	-	-	-	-	-	0.07	0.01	0.00	0.05	0.73	-

Scenario No.	Successful nsim	Truth (Diff. in RMSTs)	Method	Mean estimate (Diff. in RMSTs)	SE of mean	95% Confidence interval		Bias	Percent bias	MSE	Model SE	Empirical SE	Coverage (%)	Successful estimation (%)
						Lower	Upper							
			max MCSE	-	-	-	-	-	0.09	0.01	0.00	0.06	0.93	-
49	1828	0.087	ITT	0.140	0.030	0.083	0.198	0.053	60.51	34.42	33.76	4.24	55.91	100.00
			PP	0.115	0.030	0.056	0.175	0.028	31.81	35.85	34.87	2.01	84.08	100.00
			IPCW	0.116	0.031	0.055	0.177	0.028	32.46	36.97	35.55	2.12	84.14	100.00
			SNFTM	0.115	0.027	0.062	0.167	0.027	31.15	38.46	30.59	2.14	75.93	100.00
			min MCSE	-	-	-	-	-	0.07	0.01	0.00	0.05	0.85	-
			max MCSE	-	-	-	-	-	0.08	0.01	0.00	0.06	1.16	-
50	1768	0.064	ITT	0.109	0.057	-0.003	0.221	0.046	71.96	88.23	89.93	8.23	87.16	100.00
			PP	0.086	0.058	-0.028	0.200	0.023	35.59	90.05	91.69	5.95	93.50	100.00
			IPCW	0.086	0.076	-0.062	0.234	0.023	35.58	122.69	120.91	10.36	92.39	100.00
			SNFTM	0.089	0.053	-0.015	0.192	0.025	39.38	92.71	83.32	6.44	88.69	100.00
			min MCSE	-	-	-	-	-	0.13	0.01	0.01	0.09	0.59	-
			max MCSE	-	-	-	-	-	0.19	0.02	0.04	0.13	0.80	-
51	1774	0.095	ITT	0.148	0.055	0.040	0.256	0.053	55.21	56.87	57.94	5.98	84.50	100.00
			PP	0.120	0.056	0.011	0.229	0.025	25.87	58.60	58.76	3.90	92.84	100.00
			IPCW	0.121	0.072	-0.021	0.263	0.026	26.88	78.90	77.66	6.61	93.20	100.00
			SNFTM	0.122	0.051	0.023	0.222	0.027	28.52	60.33	53.38	4.24	87.37	100.00
			min MCSE	-	-	-	-	-	0.13	0.01	0.01	0.09	0.60	-
			max MCSE	-	-	-	-	-	0.18	0.02	0.04	0.13	0.86	-
52	1823	0.095	ITT	0.149	0.057	0.037	0.261	0.054	56.51	59.34	60.27	6.39	83.87	100.00
			PP	0.122	0.059	0.006	0.238	0.027	28.33	61.59	62.29	4.37	92.38	100.00
			IPCW	0.126	0.071	-0.013	0.265	0.031	32.15	75.95	75.47	6.47	92.20	100.00
			SNFTM	0.125	0.052	0.023	0.226	0.030	31.09	62.82	54.51	4.67	86.56	100.00
			min MCSE	-	-	-	-	-	0.13	0.01	0.01	0.09	0.62	-
			max MCSE	-	-	-	-	-	0.17	0.02	0.03	0.12	0.86	-
53	1806	0.055	ITT	0.092	0.055	-0.016	0.200	0.037	68.58	102.42	101.13	8.29	88.93	100.00
			PP	0.074	0.057	-0.038	0.186	0.019	35.50	105.81	104.69	6.79	93.24	100.00
			IPCW	0.073	0.074	-0.072	0.219	0.019	34.46	139.05	138.41	11.20	93.51	100.00
			SNFTM	0.074	0.052	-0.027	0.175	0.019	35.37	108.38	94.93	7.09	89.76	100.00
			min MCSE	-	-	-	-	-	0.13	0.01	0.01	0.09	0.58	-

Scenario No.	Successful nsim	Truth (Diff. in RMSTs)	Method	Mean estimate (Diff. in RMSTs)	SE of mean	95% Confidence interval		Bias	Percent bias	MSE	Model SE	Empirical SE	Coverage (%)	Successful estimation (%)
						Lower	Upper							
			max MCSE	-	-	-	-	-	0.18	0.02	0.04	0.13	0.74	-
54	1615	0.093	ITT	0.112	0.061	-0.008	0.232	0.019	20.33	65.25	65.76	4.35	93.25	100.00
			PP	0.107	0.068	-0.026	0.240	0.014	14.84	73.61	73.20	5.25	94.37	100.00
			IPCW	0.101	0.085	-0.065	0.267	0.008	8.57	97.63	91.76	8.94	92.57	100.00
			SNFTM	0.106	0.060	-0.012	0.223	0.013	13.72	75.62	64.44	5.50	88.73	100.00
			min MCSE	-	-	-	-	-	0.15	0.01	0.01	0.11	0.57	-
			max MCSE	-	-	-	-	-	0.23	0.03	0.03	0.16	0.79	-
55	1831	0.055	ITT	0.089	0.057	-0.023	0.201	0.034	63.03	104.51	104.53	8.13	89.84	100.00
			PP	0.072	0.060	-0.046	0.190	0.017	32.04	110.67	110.55	7.24	93.72	100.00
			IPCW	0.074	0.073	-0.068	0.216	0.019	35.18	136.64	134.17	10.86	93.77	100.00
			SNFTM	0.073	0.052	-0.030	0.176	0.018	33.62	112.27	96.30	7.49	88.91	100.00
			min MCSE	-	-	-	-	-	0.13	0.01	0.01	0.09	0.56	-
			max MCSE	-	-	-	-	-	0.17	0.02	0.03	0.12	0.73	-
56	1799	0.088	ITT	0.139	0.055	0.031	0.247	0.051	58.19	61.85	62.99	6.33	85.49	100.00
			PP	0.112	0.056	0.002	0.222	0.024	27.33	63.11	64.01	4.15	93.00	100.00
			IPCW	0.114	0.073	-0.029	0.256	0.026	29.44	85.19	84.16	7.13	93.36	100.00
			SNFTM	0.114	0.051	0.014	0.214	0.026	29.67	64.45	58.22	4.42	89.61	100.00
			min MCSE	-	-	-	-	-	0.13	0.01	0.01	0.09	0.59	-
			max MCSE	-	-	-	-	-	0.18	0.02	0.04	0.13	0.83	-

Table 7: Performance of methods across initiation non-adherence scenarios

Scenario No.	Successful nsim	Truth (Diff. in RMSTs)	Method	Mean estimate (Diff. in RMSTs)	SE of mean	95% Confidence interval		Bias	Percent bias	MSE	Model SE	Empirical SE	Coverage (%)	Successful estimation (%)
						Lower	Upper							
57	1815	0.108	ITT	0.162	0.032	0.100	0.225	0.054	49.83	3.61	29.27	29.14	59.23	100.00
			PP	0.132	0.035	0.063	0.201	0.023	21.64	1.62	32.29	32.09	89.86	100.00
			IPCW	0.131	0.037	0.058	0.204	0.023	21.08	1.79	34.41	34.78	90.58	100.00
			SNFTM	0.129	0.030	0.069	0.188	0.020	18.54	1.97	28.10	38.38	80.77	100.00
			min MCSE	-	-	-	-	-	0.07	0.01	0.00	0.05	0.69	-
			max MCSE	-	-	-	-	-	0.10	0.01	0.00	0.07	1.15	-
58	1830	0.063	ITT	0.126	0.029	0.068	0.184	0.063	99.11	7.54	46.57	45.66	44.04	100.00
			PP	0.079	0.030	0.020	0.138	0.016	24.88	1.77	47.71	46.73	91.97	100.00
			IPCW	0.080	0.032	0.018	0.142	0.017	26.30	1.97	50.01	49.18	91.53	100.00
			SNFTM	0.081	0.026	0.029	0.133	0.017	27.61	2.17	41.67	51.63	82.79	100.00
			min MCSE	-	-	-	-	-	0.07	0.00	0.00	0.05	0.64	-
			max MCSE	-	-	-	-	-	0.08	0.01	0.00	0.05	1.16	-
59	1805	0.108	ITT	0.136	0.031	0.076	0.196	0.027	25.11	1.56	28.19	28.38	85.32	100.00
			PP	0.118	0.033	0.052	0.183	0.009	8.69	1.11	30.77	30.82	94.35	100.00
			IPCW	0.119	0.035	0.050	0.188	0.011	9.78	1.32	32.66	33.46	92.96	100.00
			SNFTM	0.113	0.030	0.054	0.172	0.004	4.07	1.33	27.72	34.86	87.76	100.00
			min MCSE	-	-	-	-	-	0.07	0.00	0.00	0.05	0.54	-
			max MCSE	-	-	-	-	-	0.09	0.01	0.00	0.06	0.83	-
60	1837	0.063	ITT	0.088	0.026	0.036	0.139	0.024	38.14	2.00	41.57	41.27	85.57	100.00
			PP	0.068	0.028	0.013	0.122	0.004	6.57	1.23	44.00	43.54	94.83	100.00
			IPCW	0.069	0.029	0.011	0.126	0.005	8.21	1.38	46.15	45.89	95.05	100.00
			SNFTM	0.066	0.025	0.017	0.115	0.003	4.55	1.30	39.48	45.12	91.29	100.00
			min MCSE	-	-	-	-	-	0.06	0.00	0.00	0.04	0.51	-
			max MCSE	-	-	-	-	-	0.07	0.00	0.00	0.05	0.82	-
61	1805	0.108	ITT	0.148	0.031	0.088	0.209	0.040	36.91	2.36	28.50	28.51	73.74	100.00
			PP	0.123	0.035	0.055	0.191	0.014	13.33	1.31	32.09	32.13	93.19	100.00
			IPCW	0.122	0.037	0.050	0.193	0.013	12.37	1.42	33.72	34.01	93.52	100.00
			SNFTM	0.116	0.030	0.057	0.175	0.007	6.83	1.68	27.82	38.72	85.04	100.00
			min MCSE	-	-	-	-	-	0.07	0.00	0.00	0.05	0.58	-

Scenario No.	Successful nsim	Truth (Diff. in RMSTs)	Method	Mean estimate (Diff. in RMSTs	SE of mean	95% Confidence interval		Bias	Percent bias	MSE	Model SE	Empirical SE	Coverage (%)	Successful estimation (%)
						Lower	Upper							
			max MCSE	-	-	-	-	-	0.10	0.01	0.00	0.07	1.04	-
62	1814	0.063	ITT	0.102	0.027	0.048	0.155	0.038	60.72	3.46	43.26	42.18	72.66	100.00
			PP	0.071	0.029	0.013	0.128	0.007	11.34	1.43	46.35	46.15	93.88	100.00
			IPCW	0.071	0.031	0.011	0.132	0.008	12.60	1.59	48.67	48.54	94.38	100.00
			SNFTM	0.069	0.025	0.020	0.119	0.006	9.59	1.57	40.04	48.93	88.81	100.00
			min MCSE	-	-	-	-	-	0.06	0.00	0.00	0.04	0.54	-
			max MCSE	-	-	-	-	-	0.07	0.01	0.00	0.05	1.05	-
63	1817	0.169	ITT	0.213	0.031	0.152	0.273	0.044	26.21	1.70	18.27	18.00	70.83	100.00
			PP	0.188	0.033	0.123	0.252	0.019	11.32	0.83	19.43	19.12	91.25	100.00
			IPCW	0.188	0.035	0.119	0.257	0.020	11.69	1.00	20.96	21.40	90.53	100.00
			SNFTM	0.185	0.029	0.127	0.242	0.016	9.50	0.93	17.51	21.47	85.58	100.00
			min MCSE	-	-	-	-	-	0.07	0.00	0.00	0.05	0.66	-
			max MCSE	-	-	-	-	-	0.08	0.01	0.00	0.06	1.07	-
64	1811	0.095	ITT	0.149	0.027	0.095	0.203	0.054	57.27	3.88	28.96	28.65	49.31	100.00
			PP	0.109	0.028	0.055	0.163	0.014	15.13	0.99	29.11	28.54	91.99	100.00
			IPCW	0.110	0.029	0.053	0.166	0.015	16.07	1.10	30.49	30.12	91.44	100.00
			SNFTM	0.112	0.025	0.063	0.160	0.017	17.79	1.19	26.36	30.58	84.54	100.00
			min MCSE	-	-	-	-	-	0.06	0.00	0.00	0.04	0.64	-
			max MCSE	-	-	-	-	-	0.07	0.01	0.00	0.05	1.17	-
65	1803	0.169	ITT	0.230	0.032	0.168	0.293	0.062	36.57	2.84	18.87	18.66	51.36	100.00
			PP	0.201	0.034	0.134	0.269	0.033	19.37	1.30	20.41	19.86	84.36	100.00
			IPCW	0.200	0.037	0.128	0.272	0.031	18.51	1.41	21.87	22.19	85.14	100.00
			SNFTM	0.198	0.030	0.139	0.256	0.029	17.42	1.55	17.71	24.78	74.43	100.00
			min MCSE	-	-	-	-	-	0.07	0.01	0.00	0.05	0.84	-
			max MCSE	-	-	-	-	-	0.10	0.01	0.00	0.07	1.18	-
66	1796	0.095	ITT	0.181	0.030	0.123	0.240	0.087	91.45	8.88	31.55	31.83	18.21	100.00
			PP	0.120	0.030	0.062	0.179	0.026	26.97	1.62	31.44	31.37	87.92	100.00
			IPCW	0.121	0.031	0.061	0.182	0.027	28.15	1.77	32.79	32.78	87.31	100.00
			SNFTM	0.124	0.026	0.073	0.175	0.030	31.35	2.23	27.54	37.00	73.16	100.00
			min MCSE	-	-	-	-	-	0.07	0.00	0.00	0.05	0.77	-

Scenario No.	Successful nsim	Truth (Diff. in RMSTs)	Method	Mean estimate (Diff. in RMSTs	SE of mean	95% Confidence interval		Bias	Percent bias	MSE	Model SE	Empirical SE	Coverage (%)	Successful estimation (%)
						Lower	Upper							
			max MCSE	-	-	-	-	-	0.08	0.01	0.00	0.06	1.05	-
67	1801	0.155	ITT	0.200	0.031	0.139	0.260	0.045	28.74	1.85	19.91	19.12	69.91	100.00
			PP	0.175	0.033	0.110	0.239	0.020	12.61	0.89	21.21	20.38	91.39	100.00
			IPCW	0.176	0.035	0.106	0.245	0.021	13.35	1.10	22.83	23.08	91.12	100.00
			SNFTM	0.172	0.030	0.114	0.230	0.017	11.09	1.01	19.11	23.00	85.56	100.00
			min MCSE	-	-	-	-	-	0.07	0.00	0.00	0.05	0.66	-
			max MCSE	-	-	-	-	-	0.08	0.01	0.00	0.06	1.08	-
68	1812	0.087	ITT	0.142	0.028	0.088	0.196	0.054	62.21	4.27	31.55	31.76	48.45	100.00
			PP	0.102	0.028	0.048	0.157	0.015	16.95	1.19	31.71	32.77	91.11	100.00
			IPCW	0.103	0.029	0.046	0.160	0.016	18.17	1.31	33.27	34.20	90.78	100.00
			SNFTM	0.105	0.025	0.056	0.155	0.018	20.51	1.41	28.72	34.51	83.28	100.00
			min MCSE	-	-	-	-	-	0.07	0.00	0.00	0.05	0.67	-
			max MCSE	-	-	-	-	-	0.07	0.01	0.00	0.05	1.17	-
69	1777	0.155	ITT	0.217	0.032	0.155	0.280	0.062	40.25	3.15	20.53	20.26	49.75	100.00
			PP	0.188	0.034	0.120	0.256	0.033	21.28	1.42	22.24	21.46	84.86	100.00
			IPCW	0.188	0.037	0.116	0.260	0.033	21.21	1.58	23.81	23.87	85.31	100.00
			SNFTM	0.185	0.030	0.126	0.244	0.030	19.36	1.67	19.31	26.48	74.51	100.00
			min MCSE	-	-	-	-	-	0.07	0.01	0.00	0.05	0.84	-
			max MCSE	-	-	-	-	-	0.10	0.01	0.01	0.07	1.19	-
70	1776	0.087	ITT	0.174	0.030	0.115	0.233	0.087	99.21	9.66	34.36	34.73	18.30	100.00
			PP	0.115	0.030	0.056	0.174	0.027	31.43	1.88	34.25	34.18	84.91	100.00
			IPCW	0.116	0.031	0.055	0.177	0.029	32.82	2.07	35.72	35.90	84.63	100.00
			SNFTM	0.119	0.026	0.067	0.170	0.031	35.68	2.59	30.00	41.12	70.55	100.00
			min MCSE	-	-	-	-	-	0.07	0.00	0.00	0.05	0.85	-
			max MCSE	-	-	-	-	-	0.09	0.01	0.00	0.06	1.08	-
71	1791	0.093	ITT	0.120	0.031	0.061	0.180	0.027	29.45	1.89	32.88	34.19	83.70	100.00
			PP	0.103	0.033	0.037	0.168	0.010	10.45	1.40	35.90	37.36	92.74	100.00
			IPCW	0.104	0.036	0.035	0.174	0.011	12.05	1.68	38.21	40.71	92.35	100.00
			SNFTM	0.098	0.030	0.039	0.157	0.005	5.04	1.76	32.34	43.18	85.37	100.00
			min MCSE	-	-	-	-	-	0.08	0.00	0.00	0.05	0.61	-

Scenario No.	Successful nsim	Truth (Diff. in RMSTs)	Method	Mean estimate (Diff. in RMSTs)	SE of mean	95% Confidence interval		Bias	Percent bias	MSE	Model SE	Empirical SE	Coverage (%)	Successful estimation (%)
						Lower	Upper							
			max MCSE	-	-	-	-	-	0.09	0.01	0.00	0.07	0.87	-
72	1794	0.054	ITT	0.078	0.026	0.026	0.130	0.023	43.13	2.36	48.71	49.87	85.56	100.00
			PP	0.059	0.028	0.004	0.114	0.004	7.95	1.56	51.50	52.95	94.15	100.00
			IPCW	0.060	0.029	0.002	0.117	0.005	9.87	1.76	53.97	56.03	93.65	100.00
			SNFTM	0.058	0.025	0.009	0.107	0.003	6.19	1.70	46.23	55.54	89.97	100.00
			min MCSE	-	-	-	-	-	0.06	0.00	0.00	0.05	0.55	-
			max MCSE	-	-	-	-	-	0.07	0.00	0.00	0.05	0.83	-
73	1796	0.093	ITT	0.134	0.031	0.073	0.194	0.041	43.54	2.89	33.20	34.74	73.22	100.00
			PP	0.108	0.035	0.040	0.177	0.015	16.46	1.67	37.40	39.02	91.59	100.00
			IPCW	0.108	0.037	0.037	0.180	0.015	16.37	1.83	39.28	41.18	91.76	100.00
			SNFTM	0.101	0.030	0.041	0.160	0.007	8.01	2.10	32.49	46.84	83.02	100.00
			min MCSE	-	-	-	-	-	0.08	0.00	0.00	0.05	0.65	-
			max MCSE	-	-	-	-	-	0.10	0.01	0.00	0.07	1.04	-
74	1804	0.054	ITT	0.092	0.028	0.039	0.146	0.038	69.97	4.12	50.61	51.84	71.56	100.00
			PP	0.062	0.029	0.004	0.120	0.007	13.67	1.84	54.24	56.57	93.18	100.00
			IPCW	0.063	0.031	0.002	0.123	0.009	15.67	2.01	56.80	58.78	93.35	100.00
			SNFTM	0.061	0.025	0.011	0.110	0.006	11.40	2.02	46.84	59.93	87.20	100.00
			min MCSE	-	-	-	-	-	0.07	0.00	0.00	0.05	0.59	-
			max MCSE	-	-	-	-	-	0.08	0.01	0.00	0.05	1.06	-
75	1421	0.108	ITT	0.210	0.060	0.093	0.328	0.102	94.27	12.99	55.20	55.71	58.69	100.00
			PP	0.185	0.064	0.060	0.309	0.076	70.48	9.22	58.86	59.53	76.07	100.00
			IPCW	0.181	0.090	0.004	0.358	0.073	67.10	13.57	84.52	89.62	82.30	100.00
			SNFTM	0.180	0.057	0.068	0.292	0.072	66.31	9.02	52.72	62.74	71.01	100.00
			min MCSE	-	-	-	-	-	0.16	0.03	0.01	0.11	1.01	-
			max MCSE	-	-	-	-	-	0.26	0.05	0.05	0.18	1.31	-
76	1638	0.064	ITT	0.150	0.054	0.045	0.256	0.087	136.73	16.33	84.76	83.75	62.39	100.00
			PP	0.110	0.054	0.004	0.216	0.046	73.08	7.88	85.19	84.10	86.87	100.00
			IPCW	0.113	0.073	-0.030	0.256	0.050	77.95	12.96	117.25	119.68	88.95	100.00
			SNFTM	0.115	0.049	0.020	0.210	0.052	81.47	8.83	76.70	85.22	78.75	100.00
			min MCSE	-	-	-	-	-	0.13	0.02	0.01	0.09	0.78	-

Scenario No.	Successful nsim	Truth (Diff. in RMSTs)	Method	Mean estimate (Diff. in RMSTs)	SE of mean	95% Confidence interval		Bias	Percent bias	MSE	Model SE	Empirical SE	Coverage (%)	Successful estimation (%)
						Lower	Upper							
			max MCSE	-	-	-	-	-	0.19	0.03	0.05	0.13	1.20	-
77	1344	0.108	ITT	0.230	0.062	0.109	0.350	0.121	112.02	16.92	56.89	55.47	50.67	100.00
			PP	0.201	0.067	0.070	0.332	0.093	85.41	11.80	61.67	60.00	70.54	100.00
			IPCW	0.202	0.090	0.026	0.378	0.093	86.19	16.60	83.67	88.88	78.79	100.00
			SNFTM	0.199	0.058	0.086	0.311	0.090	83.26	12.37	53.20	67.02	61.38	100.00
			min MCSE	-	-	-	-	-	0.16	0.04	0.01	0.12	1.11	-
			max MCSE	-	-	-	-	-	0.26	0.06	0.04	0.19	1.36	-
78	1714	0.064	ITT	0.183	0.058	0.069	0.297	0.119	187.89	27.78	91.97	91.81	46.27	100.00
			PP	0.123	0.058	0.009	0.237	0.060	93.79	10.86	91.96	91.13	82.73	100.00
			IPCW	0.128	0.075	-0.019	0.275	0.064	101.39	15.98	119.64	121.95	84.89	100.00
			SNFTM	0.129	0.050	0.030	0.228	0.065	102.40	12.73	79.75	97.72	70.89	100.00
			min MCSE	-	-	-	-	-	0.14	0.02	0.01	0.10	0.87	-
			max MCSE	-	-	-	-	-	0.19	0.04	0.03	0.13	1.20	-
79	1398	0.108	ITT	0.207	0.059	0.092	0.322	0.099	91.07	12.23	54.39	54.72	60.09	100.00
			PP	0.182	0.063	0.058	0.307	0.074	68.42	8.83	58.62	58.89	77.11	100.00
			IPCW	0.180	0.088	0.007	0.353	0.071	65.85	13.24	82.62	88.82	81.60	100.00
			SNFTM	0.178	0.057	0.067	0.289	0.070	64.26	8.63	52.44	61.94	71.39	100.00
			min MCSE	-	-	-	-	-	0.16	0.03	0.01	0.11	1.04	-
			max MCSE	-	-	-	-	-	0.26	0.05	0.05	0.18	1.31	-
80	1701	0.064	ITT	0.140	0.052	0.038	0.242	0.077	120.46	13.47	82.09	81.81	68.96	100.00
			PP	0.107	0.053	0.002	0.211	0.043	67.67	7.28	84.24	82.95	88.07	100.00
			IPCW	0.109	0.072	-0.032	0.250	0.045	71.53	12.19	115.61	118.61	89.18	100.00
			SNFTM	0.110	0.048	0.017	0.203	0.046	73.12	7.89	75.42	84.15	81.25	100.00
			min MCSE	-	-	-	-	-	0.13	0.01	0.01	0.09	0.76	-
			max MCSE	-	-	-	-	-	0.18	0.03	0.05	0.13	1.12	-
81	1471	0.108	ITT	0.227	0.058	0.113	0.340	0.118	109.24	15.96	53.52	52.94	47.31	100.00
			PP	0.195	0.064	0.070	0.320	0.087	80.07	10.52	58.89	57.44	71.65	100.00
			IPCW	0.193	0.083	0.029	0.356	0.084	77.95	13.50	77.73	79.97	79.95	100.00
			SNFTM	0.191	0.055	0.083	0.298	0.082	76.02	10.32	50.67	61.21	64.17	100.00
			min MCSE	-	-	-	-	-	0.15	0.03	0.01	0.11	1.04	-

Scenario No.	Successful nsim	Truth (Diff. in RMSTs)	Method	Mean estimate (Diff. in RMSTs)	SE of mean	95% Confidence interval		Bias	Percent bias	MSE	Model SE	Empirical SE	Coverage (%)	Successful estimation (%)
						Lower	Upper							
			max MCSE	-	-	-	-	-	0.23	0.05	0.03	0.16	1.30	-
82	1715	0.064	ITT	0.168	0.056	0.059	0.277	0.105	165.09	22.19	87.68	87.61	52.24	100.00
			PP	0.116	0.057	0.005	0.228	0.053	83.14	9.36	90.08	88.46	85.66	100.00
			IPCW	0.120	0.073	-0.024	0.263	0.056	88.39	13.85	116.82	118.28	86.60	100.00
			SNFTM	0.120	0.049	0.024	0.217	0.057	89.57	10.34	77.74	90.85	75.80	100.00
			min MCSE	-	-	-	-	-	0.13	0.02	0.01	0.10	0.82	-
			max MCSE	-	-	-	-	-	0.18	0.03	0.03	0.13	1.21	-
83	1313	0.155	ITT	0.127	0.060	0.009	0.245	- 0.028	-17.85	3.07	38.85	40.79	90.94	100.00
			PP	0.105	0.065	-0.022	0.232	- 0.050	-32.28	4.51	42.00	43.19	86.75	100.00
			IPCW	0.110	0.092	-0.071	0.291	- 0.045	-28.96	7.49	60.33	63.20	90.47	100.00
			SNFTM	0.103	0.059	-0.012	0.218	- 0.052	-33.55	4.85	37.80	44.77	80.50	100.00
			min MCSE	-	-	-	-	-	0.17	0.02	0.01	0.12	0.79	-
			max MCSE	-	-	-	-	-	0.27	0.05	0.05	0.19	1.09	-
84	1680	0.088	ITT	0.091	0.054	-0.015	0.197	0.003	3.94	3.17	61.70	59.96	95.00	100.00
			PP	0.061	0.055	-0.047	0.170	- 0.026	-30.04	4.13	63.19	61.62	91.96	100.00
			IPCW	0.066	0.076	-0.082	0.214	- 0.022	-25.05	7.51	90.32	89.02	92.83	100.00
			SNFTM	0.063	0.050	-0.034	0.161	- 0.024	-27.71	4.12	56.84	62.65	89.23	100.00
			min MCSE	-	-	-	-	-	0.13	0.01	0.01	0.09	0.53	-
			max MCSE	-	-	-	-	-	0.19	0.03	0.28	0.14	0.76	-
85	1416	0.155	ITT	0.150	0.059	0.035	0.265	- 0.005	-3.43	2.32	37.95	38.51	94.99	100.00
			PP	0.118	0.065	-0.009	0.246	- 0.037	-23.60	3.71	42.01	42.82	92.02	100.00
			IPCW	0.127	0.087	-0.043	0.297	- 0.028	-17.91	5.66	56.38	57.70	93.01	100.00
			SNFTM	0.117	0.057	0.006	0.227	- 0.039	-24.83	4.26	36.53	46.18	83.12	100.00
			min MCSE	-	-	-	-	-	0.16	0.01	0.01	0.11	0.58	-
			max MCSE	-	-	-	-	-	0.24	0.03	0.03	0.17	1.00	-
86	1715	0.088	ITT	0.117	0.058	0.004	0.230	0.029	32.90	4.77	65.76	65.94	91.08	100.00
			PP	0.069	0.059	-0.046	0.185	- 0.019	-21.28	4.15	67.38	65.40	93.29	100.00
			IPCW	0.075	0.076	-0.075	0.225	- 0.013	-14.41	7.11	88.28	88.87	93.46	100.00
			SNFTM	0.072	0.051	-0.028	0.172	- 0.015	-17.62	4.28	58.33	67.54	88.80	100.00
			min MCSE	-	-	-	-	-	0.14	0.01	0.01	0.10	0.60	-

Scenario No.	Successful nsim	Truth (Diff. in RMSTs)	Method	Mean estimate (Diff. in RMSTs)	SE of mean	95% Confidence interval		Bias	Percent bias	MSE	Model SE	Empirical SE	Coverage (%)	Successful estimation (%)
						Lower	Upper							
			max MCSE	-	-	-	-	-	0.19	0.02	0.03	0.13	0.76	-
87	1415	0.093	ITT	0.192	0.059	0.076	0.307	0.098	105.75	14.32	63.36	64.81	59.51	100.00
			PP	0.168	0.064	0.043	0.292	0.075	80.27	10.45	68.31	69.15	77.24	100.00
			IPCW	0.170	0.090	-0.006	0.346	0.077	82.73	17.15	97.71	107.61	81.30	100.00
			SNFTM	0.164	0.057	0.053	0.276	0.071	76.34	10.11	61.23	70.93	73.29	100.00
			min MCSE	-	-	-	-	-	0.16	0.03	0.01	0.11	1.04	-
			max MCSE	-	-	-	-	-	0.27	0.06	0.05	0.19	1.30	-
88	1696	0.055	ITT	0.134	0.052	0.031	0.236	0.079	144.81	16.26	95.94	93.99	67.57	100.00
			PP	0.100	0.053	-0.005	0.205	0.046	83.39	8.82	98.43	96.00	87.32	100.00
			IPCW	0.103	0.073	-0.039	0.245	0.048	88.60	14.93	136.37	139.73	88.72	100.00
			SNFTM	0.104	0.048	0.010	0.198	0.049	90.09	9.75	88.21	98.76	79.48	100.00
			min MCSE	-	-	-	-	-	0.12	0.02	0.01	0.09	0.77	-
			max MCSE	-	-	-	-	-	0.19	0.03	0.07	0.13	1.14	-
89	1461	0.093	ITT	0.210	0.060	0.092	0.328	0.117	125.65	18.83	64.83	66.60	50.65	100.00
			PP	0.179	0.066	0.049	0.309	0.086	92.41	12.98	71.35	73.49	73.03	100.00
			IPCW	0.181	0.089	0.007	0.356	0.088	94.85	17.72	96.69	100.20	80.08	100.00
			SNFTM	0.174	0.057	0.062	0.287	0.081	87.31	12.89	61.50	78.87	65.23	100.00
			min MCSE	-	-	-	-	-	0.16	0.04	0.01	0.11	1.04	-
			max MCSE	-	-	-	-	-	0.24	0.06	0.04	0.17	1.31	-
90	1778	0.055	ITT	0.164	0.056	0.054	0.273	0.109	199.64	27.52	102.59	102.85	49.72	100.00
			PP	0.110	0.057	-0.002	0.223	0.056	102.37	11.58	105.29	103.65	84.25	100.00
			IPCW	0.115	0.074	-0.030	0.259	0.060	110.13	16.95	137.57	137.61	87.01	100.00
			SNFTM	0.114	0.049	0.017	0.210	0.059	108.28	12.66	90.82	107.13	72.48	100.00
			min MCSE	-	-	-	-	-	0.13	0.02	0.01	0.09	0.80	-
			max MCSE	-	-	-	-	-	0.18	0.03	0.04	0.13	1.19	-

Supplementary Appendix E: Performance of methods - additional measures

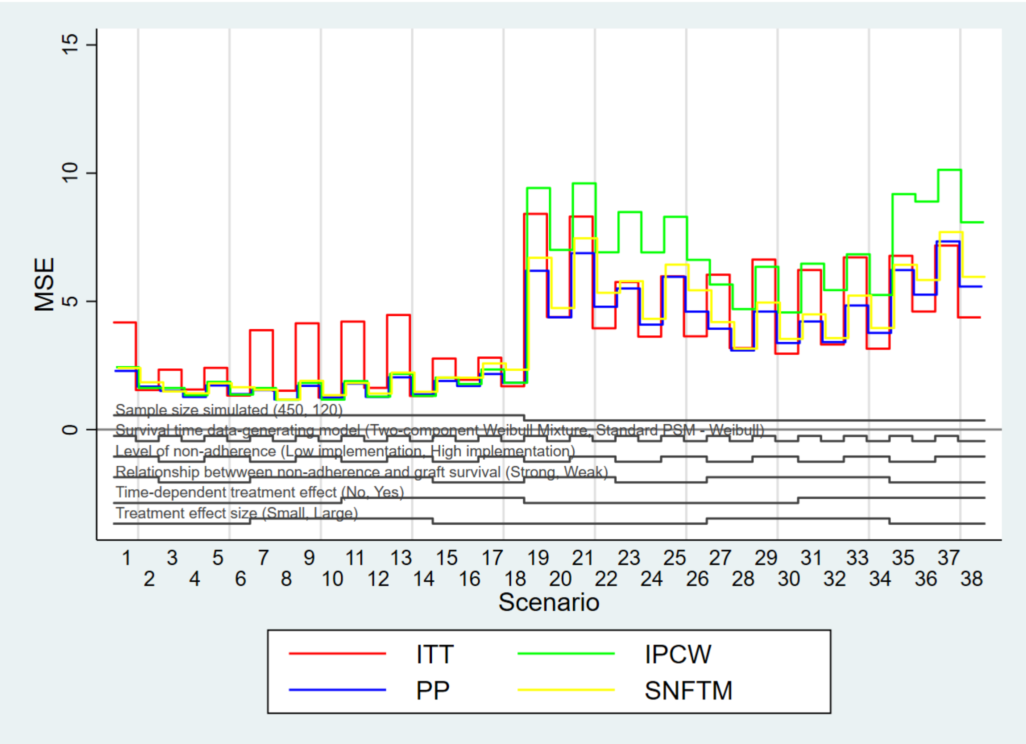
1.Performance of methods across implementation non-adherence scenarios

1.1 Mean square error

The performance of methods based on mean square error (MSE) is presented in Figure 1. MSE is a useful performance measure because it combines both bias and variability in a single measure. The methods' performance in terms of MSE is presented as a percent of the truth. PP and g-methods produced the lowest MSE across scenarios with an average MSE of 1.7% for large sample size and this remained below 3% in most of these scenarios. In contrast, ITT produced much higher MSE with higher values up to 4.5% in some cases.

Interestingly, ITT showed improved MSE performance in scenarios with a smaller sample size (Scenarios 19-38) with g-methods and PP lagging behind in these scenarios. There is some fluctuation in MSE performance resulting from a combination of treatment effect size with other factors such as the level of non-adherence and the relationship between non-adherence and graft survival outcome. PP analysis produced good MSE performance compared to the alternative methods, although the method struggled to cope in scenarios with a small sample size, as did g-methods

Figure 19: MSE in the estimation of the difference in RMSTs across implementation non-adherence scenarios

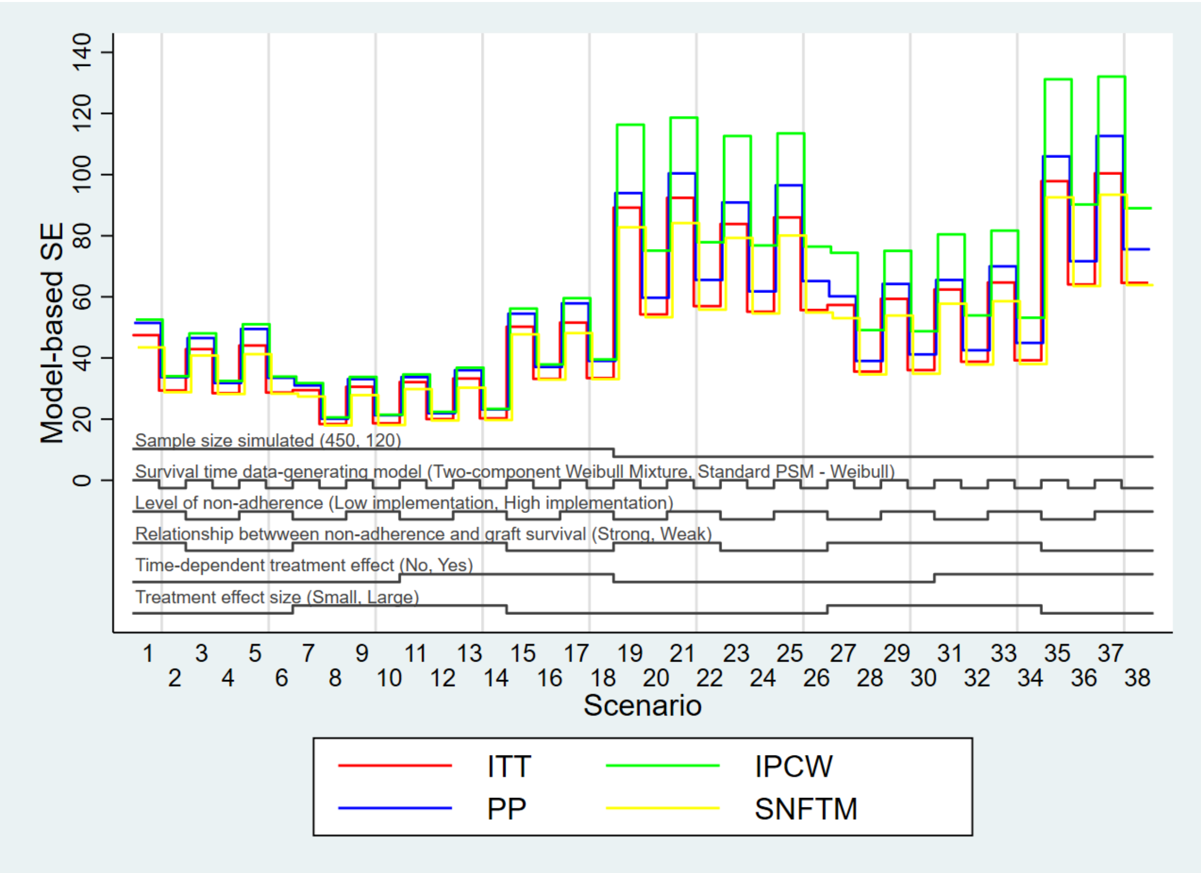


1.2 Model-based standard error

The performance of adjustment methods in terms of model-based standard error (ModSE) across the implementation non-adherence scenarios is presented in Figure 2. These figures are reported as a percentage of the truth to aid the comparisons of methods across the range of scenarios. As illustrated in the nested loop plot (Figure 2), scenarios with a smaller sample size (19-38) produced higher ModSE.

For ModSE, SNFTM performed the best in all the implementation non-adherence scenarios with IPCW and PP showing lower performance across scenarios. ITT always produced ModSE higher than SNFTM but better than IPCW and PP in most cases with higher ModSE % in scenarios with a small sample size. In addition to sample size, treatment effect size, time-dependent treatment effect and the relationship between non-adherence and graft survival outcome are the factors that had a noticeable influence on ModSE performance. A larger treatment effect size (Scenario 7-14 and 27-34) contributed to generating higher ModSE across methods, although SNFTM remained the best performing method in all of these scenarios.

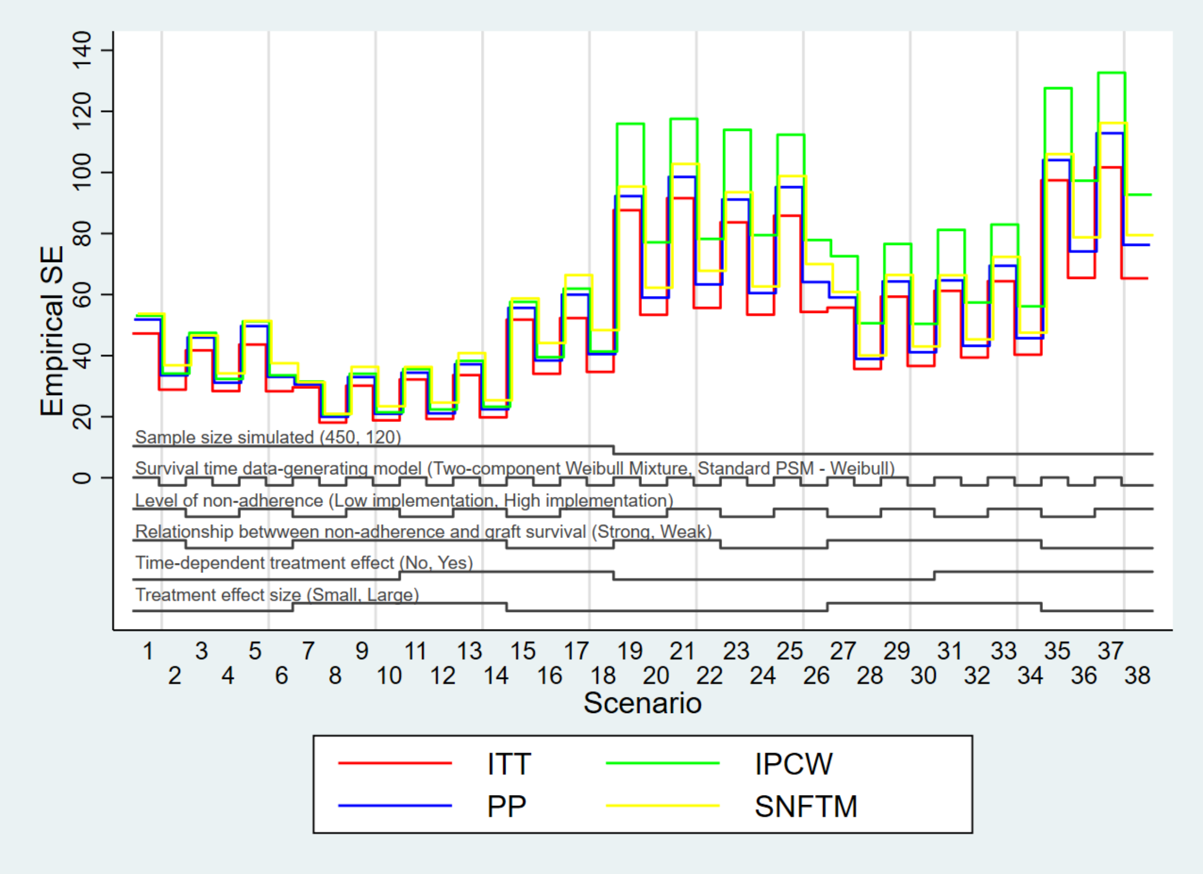
Figure 20: Model-based SE in the estimation of the difference in RMSTs across implementation non-adherence scenarios



1.3 Empirical standard error

The empirical standard error (EmpSE) represents the standard deviation of the estimates across the successful simulations (out of 1900 iterations). The performance of methods in terms of EmpSE across the implementation scenarios is illustrated in Figure 3. The results favoured ITT across scenarios, however, this should be interpreted with caution because ITT has generated the highest bias and produced the worst ModSE performance across all implementation non-adherence scenarios. Nevertheless, EmpSE performance is reported to provide the full picture in terms of the simulation results and to comply with the study protocol.

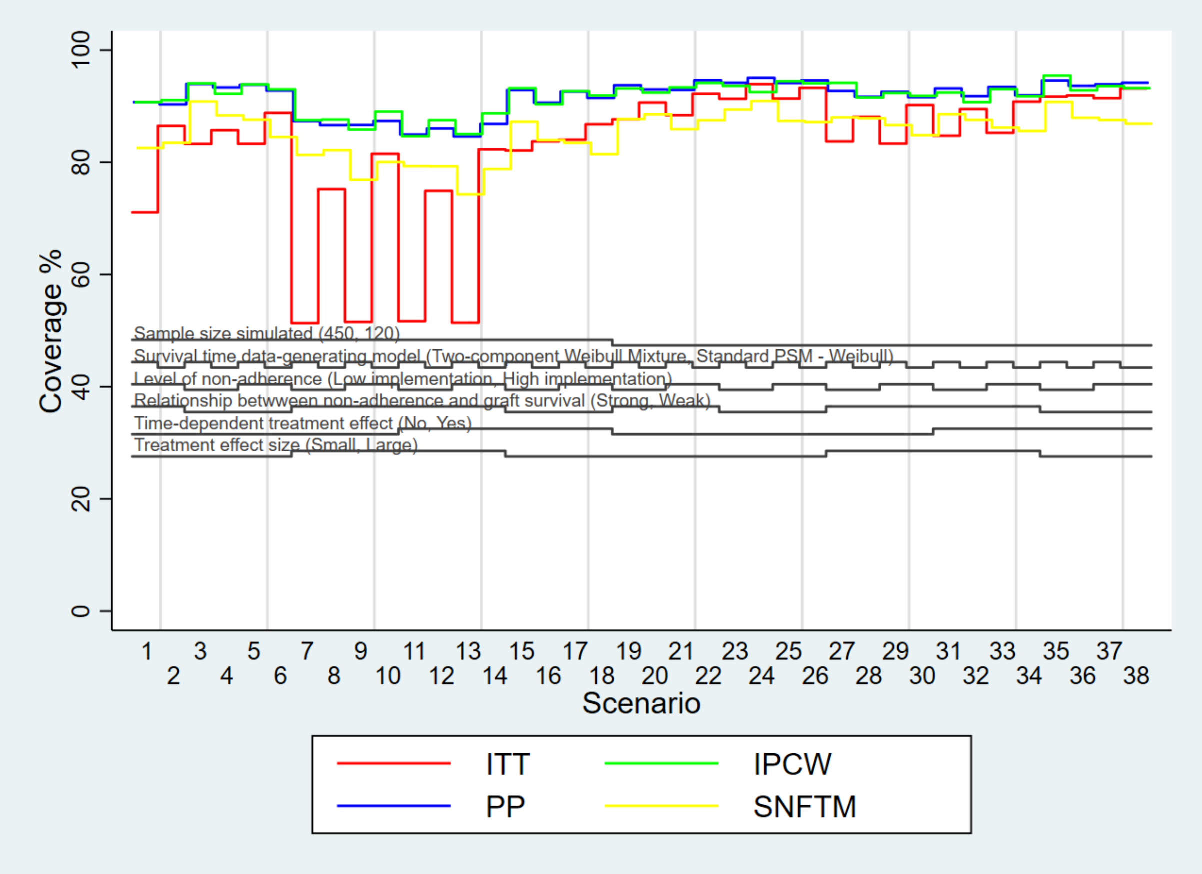
Figure 21: Empirical SE in the estimation of the difference in RMSTs across implementation non-adherence scenarios



1.4 Coverage

Figure 4 reports coverage percent for each method across 38 implementation non-adherence scenarios. Coverage percent represents the proportion of times that the 95% confidence interval contains the true value of the estimated parameter (i.e., the difference in RMST). Coverage performance data favoured IPCW and PP across scenarios with coverage percent reaching more than 90% in most cases. In contrast, ITT produced lower coverage especially in scenarios with large sample size. SNFTM struggled in terms of coverage in scenarios with small sample size.

Figure 22: Coverage in the estimation of the difference in RMSTs across implementation non-adherence scenarios



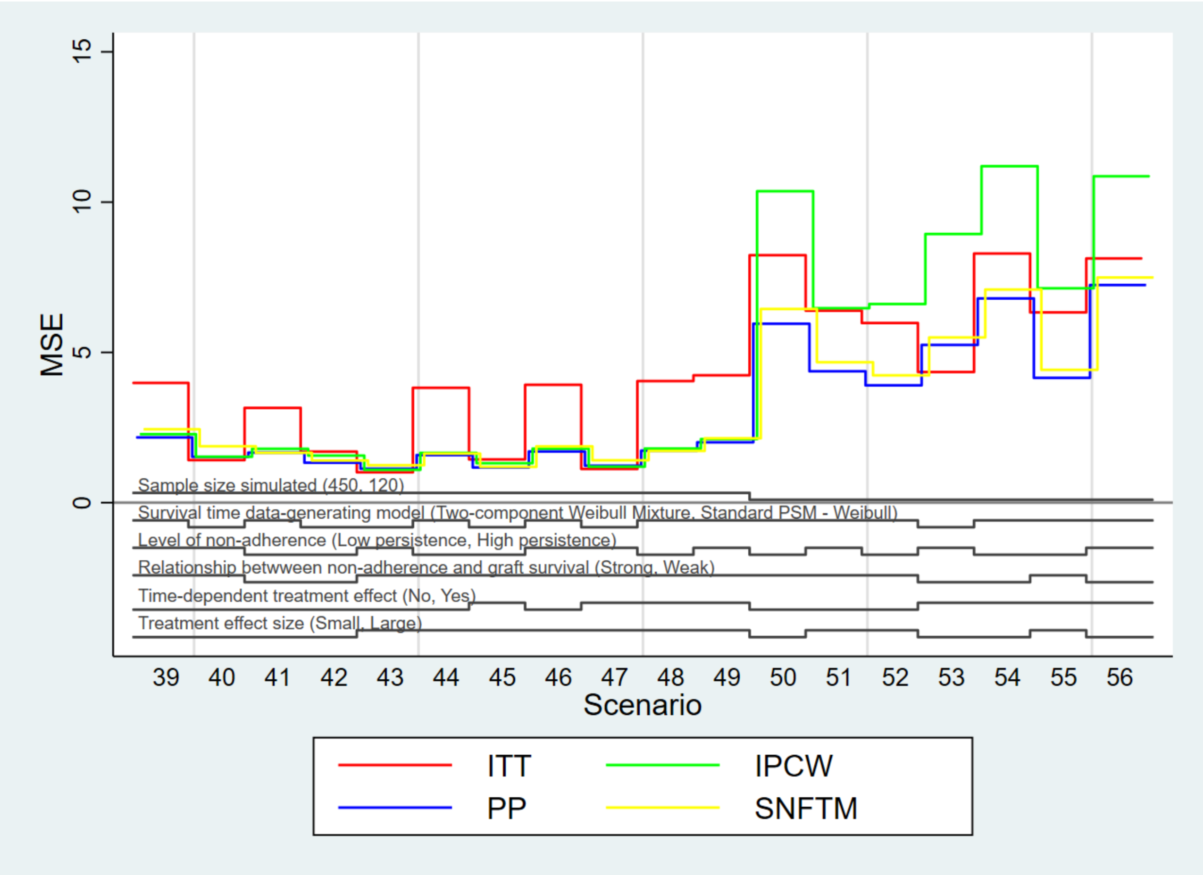
2.Performance of methods across persistence non-adherence scenarios

2.1 Mean square error

Figure 5 illustrates the results of methods performance in terms of MSE. The values are expressed as a percentage of the truth to aid comparison using the nested loop plots. As shown in the figure, PP permed best especially in scenarios with a small sample size. The IPCW performance was influenced by the combination of sample size and treatment effect size with a small sample size and low persistence leading to a higher MSE percentage.

SNFTM, IPCW and PP generally did well in scenarios with a large sample size (n=450) with MSE below 2.5 % in most scenarios. In these scenarios, ITT produced higher MSE with clear fluctuation influenced by survival time DGM and level of non-adherence. MSE performance results were influenced by sample size, with IPCW performing worst in scenarios with a small sample size with MSE values reaching up to 10 % of the truth. PP performed better than ITT in most scenarios (See Figure 5)

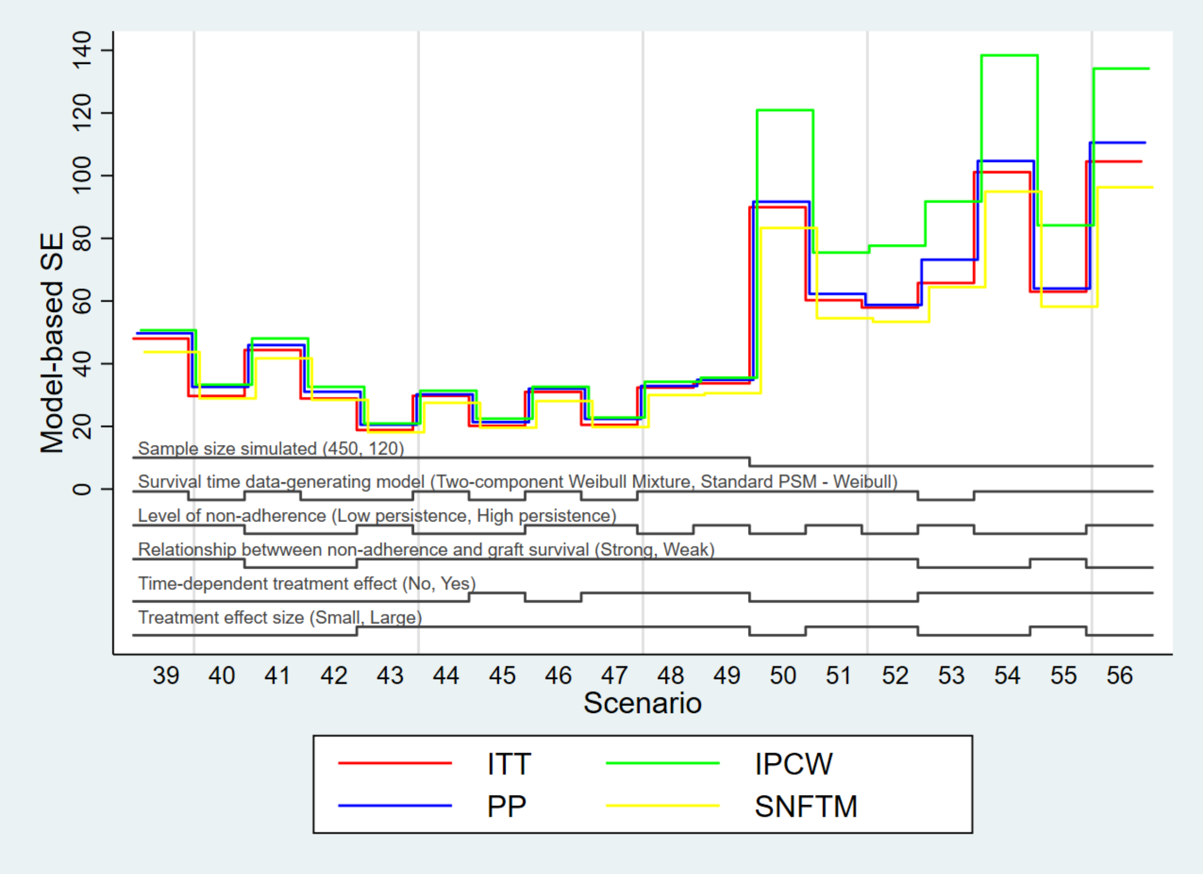
Figure 23: MSE in the estimation of the difference in RMSTs across persistence non-adherence scenarios



2.2 Model-based standard error

Figure 6 shows the performance results in adjusting for persistence non-adherence using ModSE. SNFTM with g-estimation performed the best across all persistence non-adherence scenarios with the most noticeable trend relating to sample size. In scenarios with a large sample size of 450 observations (39-49), SNFTM performed better than scenarios with a small size of 120 (50-56). The impact of a small sample size on ModSE was much bigger for IPCW with substantially higher values in scenarios with a small sample size. Large treatment effect size (in combination with small sample size) led to an even higher ModSE percentage for IPCW.

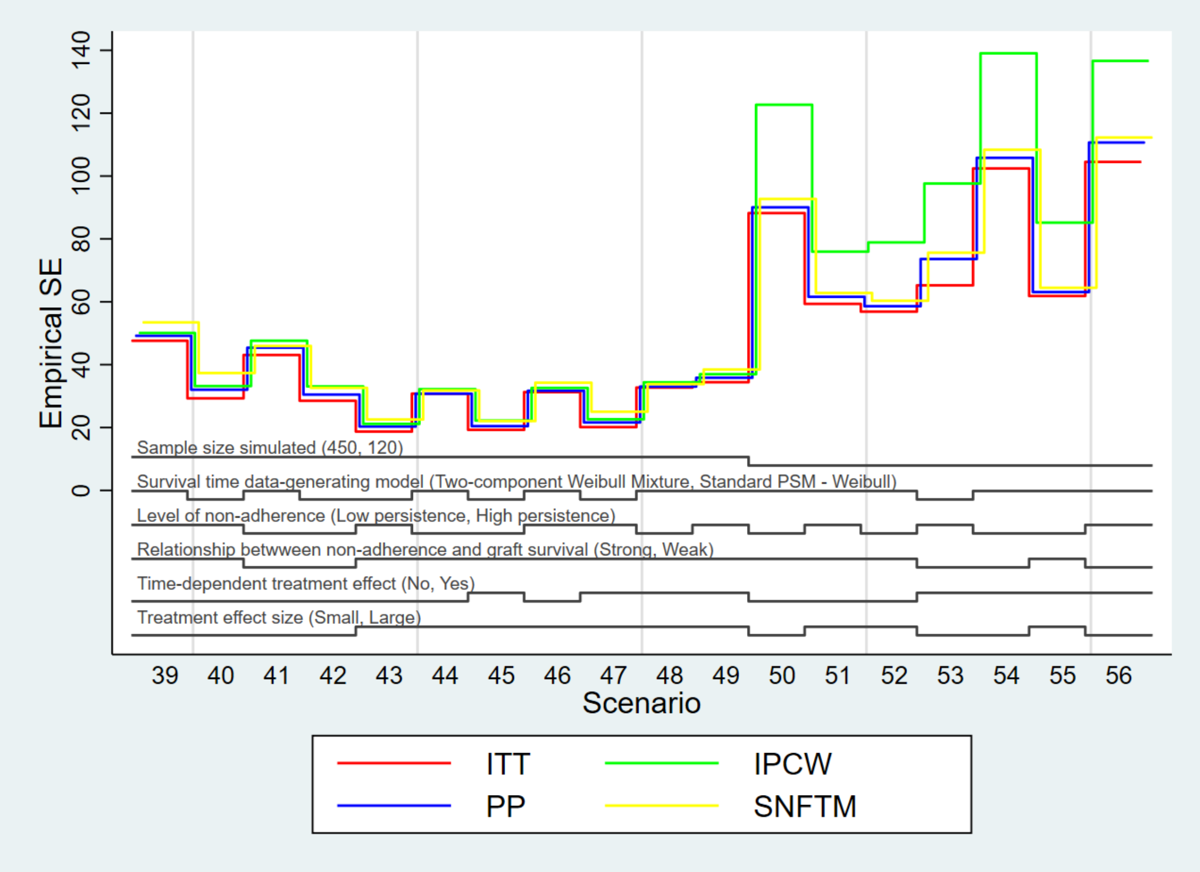
Figure 24: Model-based SE in the estimation of the difference in RMSTs across persistence non-adherence scenarios



2.3 Empirical standard error

Figure 7 illustrates methods performance using EmpSE across the 18 persistence non-adherence scenarios (39-56). These results show ITT as the best performing method followed by PP with SNFTM and IPCW produced higher EmpSE. These results should be interpreted with caution as discussed in section 1.3

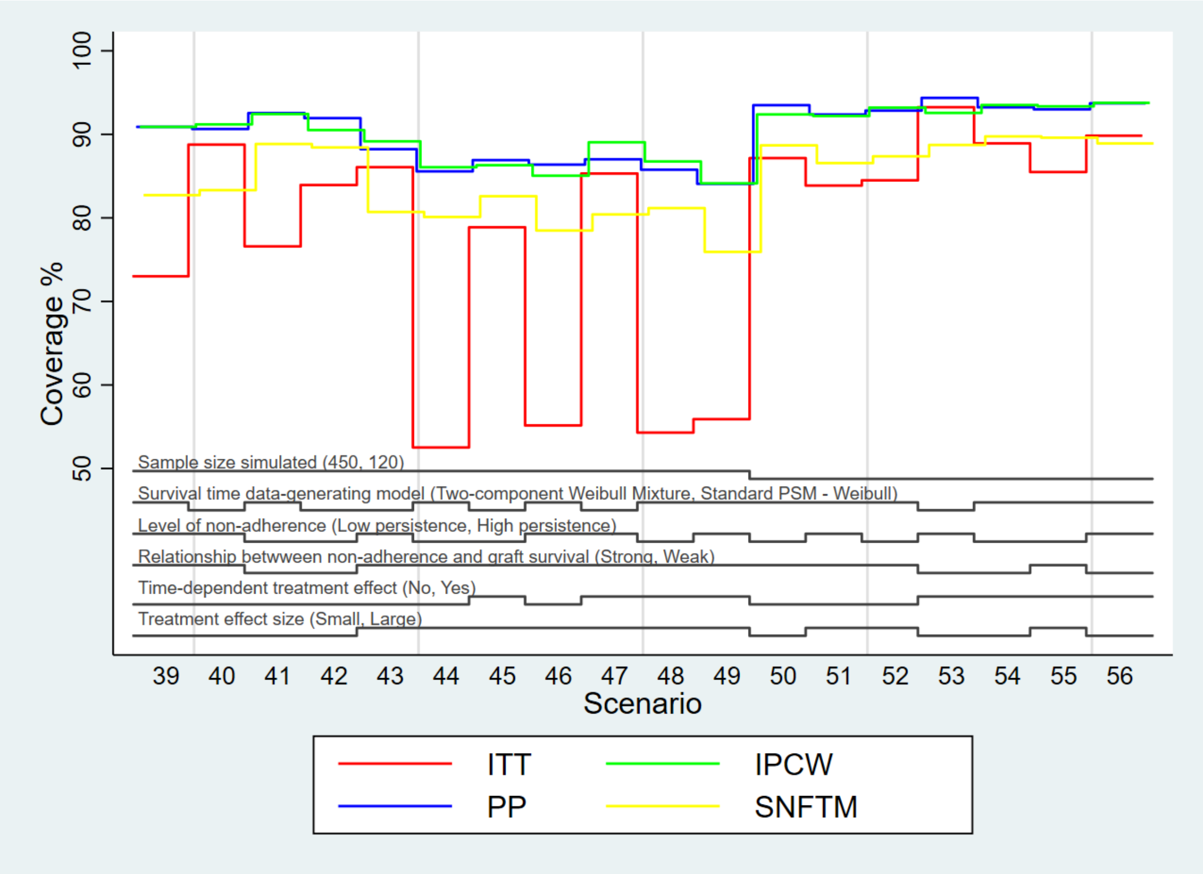
Figure 25: Empirical SE in the estimation of the difference in RMSTs across persistence non-adherence scenarios



2.4 Coverage

Figure 8 presents coverage percent showing the probability that the 95% confidence interval contains the true value of the estimated parameter (i.e., the difference in RMST) across the successful simulations. IPCW performed best in most scenarios but PP did better in scenarios with large treatment effect size. IPCW and PP coverage was generally very high reaching more than 90% in most cases. SNFTM lagged behind IPCW and PP in terms of coverage, but the method still produced better performance compared to ITT, which was the worst across all scenarios. Coverage was affected by the DGM with ITT resulted in coverage lower than 60% in some cases.

Figure 26: Coverage in the estimation of the difference in RMSTs across persistence non-adherence scenarios

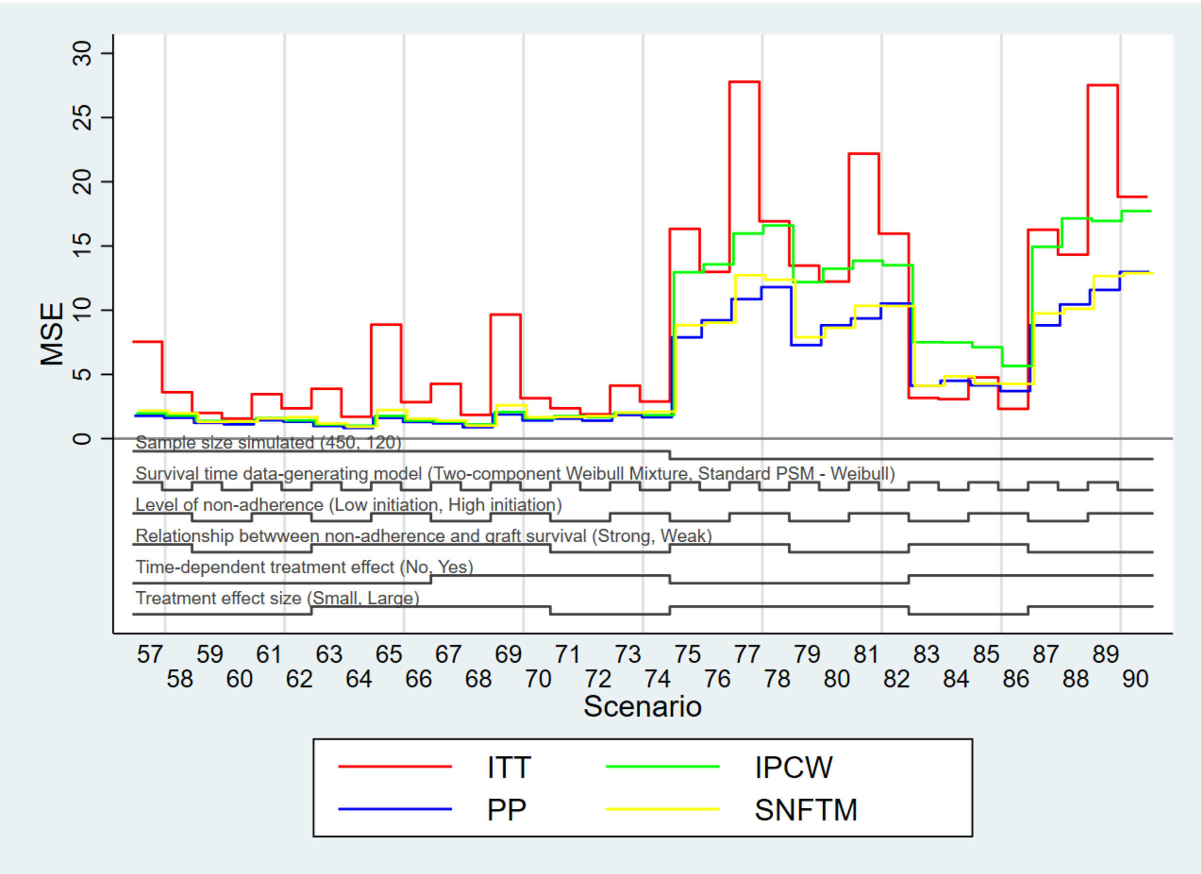


3.Performance of methods across initiation non-adherence scenarios

3.1 Mean square error

Figure 9 shows the results using MSE as a performance measure. G-methods (SNFTM and IPCW) resulted in very low MSE (expressed as parentage of the true difference in RMST) compared to ITT across scenarios with a large sample size. For these methods, MSE percent was less than 9.7% with an average of 2.1% across compared with an average MSE of 3.7% for ITT across the same scenarios. G-methods struggled in terms of MSE performance in scenarios with a small sample size combined with high treatment effect size as a key contributing factor. ITT was the worst-performing method across most scenarios with an MSE of more than 25% in some scenarios with a small sample size. ITT did better in scenarios with a small sample size, small treatment effect size and time-dependent treatment effect (Scenarios 83-86). The influence of each factor combined with other factors specified in the simulation study design is illustrated by the descriptors in the nested loop plots (see Figure 9).

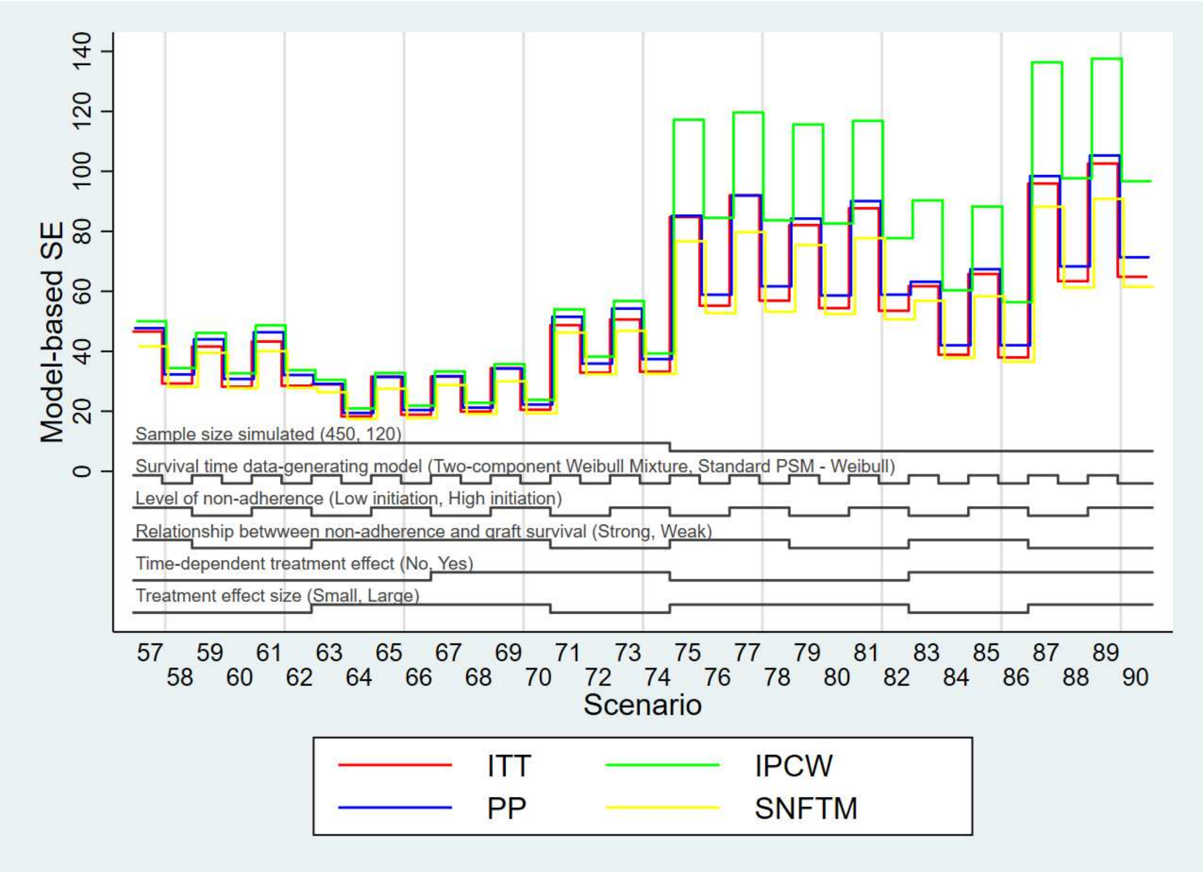
Figure 27: MSE in the estimation of the difference in RMSTs across initiation non-adherence scenarios



3.2 Model-based standard error

Methods performance based on ModSE (as a percentage of the truth) is presented in Figure 10. SNFTM with g-estimation showed better performance compared to the alternative methods across all 34 initiation non-adherence scenarios. ModSE in scenarios with a large sample size (n=450) is better than scenarios with a small sample size as illustrated in the nested loop plot (Figure 10). IPCW produced higher ModSE in scenarios with a small sample size (Scenarios 75-90). Similar to implementation and persistence non-adherence, treatment effect size had a large influence on methods performance when it comes to adjusting treatment effect for initiation non-adherence.

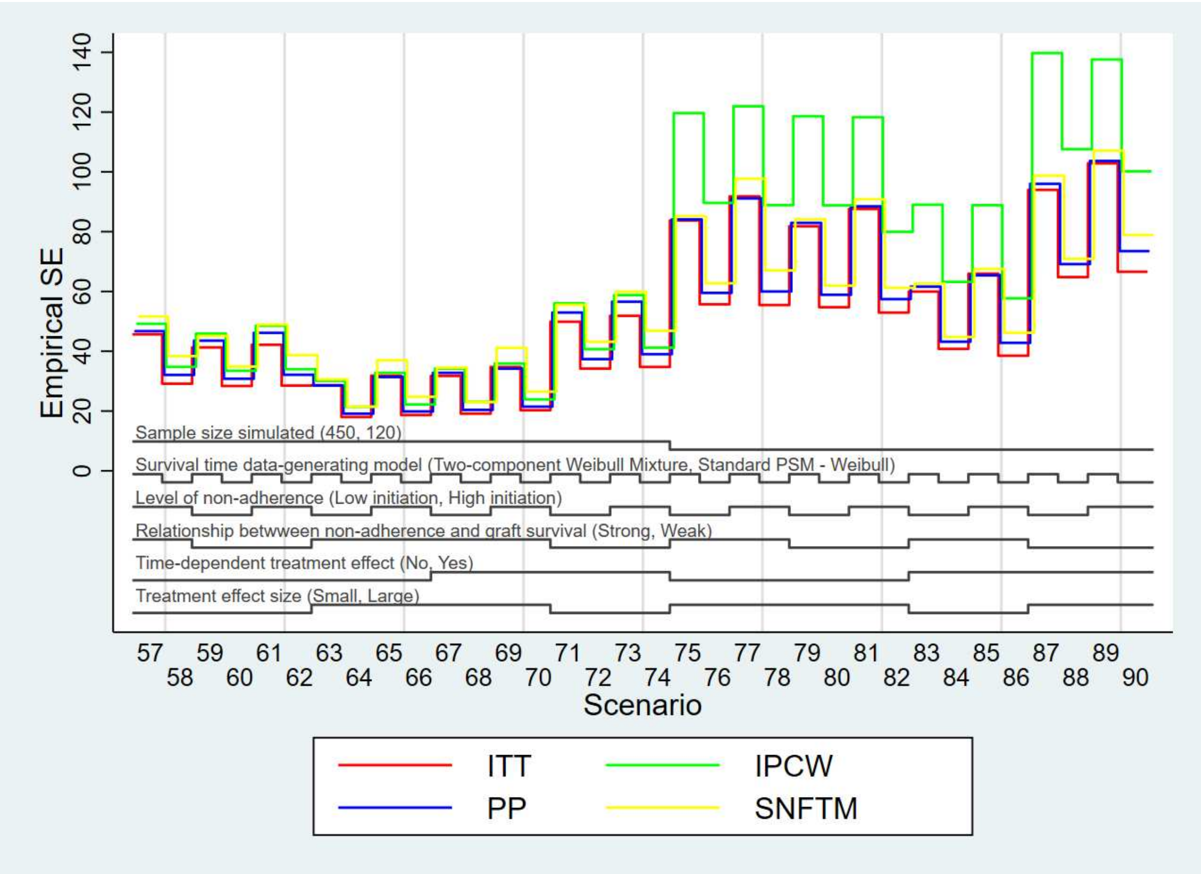
Figure 28: Model-based SE in the estimation of the difference in RMSTs across initiation non-adherence scenarios



3.3 Empirical standard error

Figure 11 illustrates the performance of methods using EmpSE across 34 initiation non-adherence scenarios. As shown, ITT and PP produced better results compared to g-methods with IPCW generated higher EmpSE in scenarios with a small sample size. As discussed earlier, the EmpSE results should be interpreted with caution and read alongside the results of the other performance measures.

Figure 29: Empirical SE in the estimation of the difference in RMSTs across initiation non-adherence scenarios



3.4 Coverage

Figure 12 reports coverage percent generated by each method across the 34 initiation non-adherence scenarios. IPCW performed best across 21 out of 34 scenarios with a very high coverage percentage up to 95% (87.2 % on average). This was followed by PP and then SNFTM with ITT produced the worst coverage across scenarios. The level of non-adherence was a noticeable contribution to the magnitude of coverage for ITT in particular lower levels of initiation non-adherence leading to lower coverage percentage.

Figure 30: Coverage in the estimation of the difference in RMSTs across initiation non-adherence scenarios