

**PROPERTIES OF INVERSE PROBABILITY OF ADHERENCE
WEIGHTED ESTIMATOR OF THE PER-PROTOCOL EFFECT FOR
SUSTAINED TREATMENT STRATEGIES UNDER DIFFERENT
DATA-GENERATING MECHANISMS AND ADHERENCE PATTERNS**

LUCY MOSQUERA

*Department of Statistics, University of British Columbia, 3182 Earth Sciences Building, 2207 Main Mall,
Vancouver, BC Canada V6T 1Z4, Canada.
Email: bellemarelucy@gmail.com*

MOHAMMAD EHSANUL KARIM*

*School of Population and Public Health, University of British Columbia, 2206 East Mall, Vancouver, BC, V6T
1Z3, Canada; Centre for Health Evaluation and Outcome Sciences, 588 - 1081 Burrard Street; St. Paul's
Hospital, Vancouver, BC, V6Z 1Y6, Canada.
Email: ehsan.karim@ubc.ca*

MD. BELAL HOSSAIN

*School of Population and Public Health, University of British Columbia, 2206 East Mall, Vancouver, BC, V6T
1Z3, Canada.
Email: belal.hossain@ubc.ca*

SUMMARY

Inverse Probability (of Adherence) Weighted per-protocol (IPW-PP) estimators are getting popular in addressing medication non-adherence while analyzing pragmatic trial data. However, their finite sample properties under different data generating mechanisms (DGMs) have not been investigated comprehensively. In the current work, we investigated the finite sample performances of such estimators in the context of a pragmatic randomized controlled trial. We compared the performances of IPW-PP estimators with commonly used naive and baseline-adjusted per-protocol estimators, under different DGMs emulating pragmatic trials, comparing two sustained treatment strategies, possibly with a non-null effect. DGMs include (i) different roles of a baseline variable; whether future time-varying prognostic factors are impacted by past adherence; and whether the baseline variable is measured, (ii) whether adherence patterns observed in two arms are differential, and when we have access to measurements of adherence and confounders that are recorded infrequently (sparsely). When baseline confounders are adjusted, we generally obtain unbiased estimates, but if some necessary variables are not measured, the IPW-PP estimator may still be preferable. High non-adherence patterns might negatively impact IPW-PP effect estimators, particularly when DGMs include confounding that may be influenced by previous adherence history. We used the above estimators to analyze a case study from the Lipid Research Clinics Coronary Primary Prevention Trial data in the presence of non-adherence.

Keywords and phrases: Non-adherence, causal inference, pragmatic trials, per-protocol

AMS Classification: 62D20

* Corresponding author, equal contribution as first author.

© Institute of Statistical Research and Training (ISRT), University of Dhaka, Dhaka 1000, Bangladesh.

1 Introduction

Treatment effect estimation in pragmatic trials

Clinical trials are conducted to obtain regulatory approval for a new drug or medical device. Common features of clinical trials include restrictive eligibility criteria, delivery of treatment by clinicians who are experts in the area, and stringent monitoring and follow-up. The pragmatic trial, on the other hand, aims to demonstrate the effectiveness of a treatment strategy in a broader range of patient groups under realistic clinical care scenarios (Schwartz and Lellouch, 1967; Zwarenstein et al., 2008). These scenarios may include mimicking settings encountered in regular clinical practice: comparing the effectiveness of a new treatment with that of an active comparator (e.g., usual care), unblinded treatment assignment, allowing patients to deviate from the protocol (e.g., non-adherence to the assigned medication), inclusion of heterogeneous subjects, prioritizing patient-centric outcomes and adequate follow-up times (Gamerman et al., 2019). Particularly for diseases that require sustained treatment strategies, pragmatic trials can inform and guide healthcare professionals, patients, and other stakeholders about the implication of taking that treatment in everyday practice.

Inadequacy of intention-to-treat analysis, and proposed alternatives

Intention-to-treat (ITT) analysis, usually done in clinical trials, compares the risk of the outcome in two treatment groups and provides asymptotically consistent effect estimates of randomization on participants' outcomes (Hernán and Hernández-Díaz, 2012). That means, ITT estimates the effect of randomized to a treatment group, irrespective of whether the subject adhered to the assigned treatment strategies during the follow-up. But for pragmatic trials, where deviation from protocol (e.g., partial adherence to medication under real-world clinical condition) is possible, such an estimate may have less value for a patient or a caregiver deciding about treatment. They may prefer to know the effect of the treatment strategies on the subjects who followed the protocol. As-treated and per-protocol are two analysis strategies that are often proposed to estimate the effect of receiving treatment in trials with incomplete adherence. In a naive as-treated analysis, person-time contributed by a subject under a given treatment are considered whether a subject switches or stops taking a treatment to calculate the as-treated effect. In a naive per-protocol analysis, subjects who deviate from the treatment protocol they were randomized to are either censored at the time of protocol deviation, or excluded from the analysis. Unfortunately, none of these naive analysis strategies will generally produce unbiased treatment effect estimates unless the non-adherence happens completely at random (Hernán and Hernández-Díaz, 2012).

Estimators of the per-protocol effect

For naive per-protocol analyses, subjects are artificially censored at the time when they first stop or switch medication or is lost to follow-up. It is reasonable to expect that patients with poor health outcomes would tend to switch or stop medication, and hence the resulting censoring may not be independent. Therefore, selection bias induced by dependent censoring creating during the naive per-protocol analysis biases the results. To correct the estimates, pre- and post-treatment prognostic

factors, that confounds the relationship between adherence and the study outcome, need to be measured and adjusted. It has been shown that, the inverse probability of (adherence) weighted (IPW) Kaplan-Meier and Cox partial likelihood estimators will provide asymptotically consistent estimates of the per-protocol effect, if the model is correctly specified (Robins and Finkelstein, 2000). Adherence to medication pattern may also impact some of these time-dependent confounders, and IPW marginal structural model were subsequently proposed to reduce bias due to treatment-confounder feedback (Hernán and Robins, 2017; Murray et al., 2021). Similarly, as-treated analyses can also be vulnerable to post-baseline time-varying confounding that may impact future adherence, and can be corrected via IP treatment weighting, with certain assumptions (Danaei et al., 2013). Despite appropriate methods developed in the statistical literature, most practitioners were reluctant to apply anything other than ITT. For example, the landmark Coronary Drug Project (CDP) highlighted the difficulty of working with per-protocol estimates in the presence of non-adherence (Group, 1980), discouraging practitioners over the years to work with per-protocol effect estimates. A pair of recent papers by Murray and Hernán (Murray and Hernán, 2016, 2018) have recently reanalyzed the original CDP data, and illustrated the benefits of using IPW estimators of the per-protocol effect. These illustrations renewed the interest in this estimator in various subsequent applications (Lu et al., 2019; Wanis et al., 2020; Murray et al., 2020) and methodological recommendations (Hernán and Robins, 2017; Mo et al., 2020; Murray et al., 2019).

Gap in the literature

Almost all of the previous simulation studies designed to assess the properties of the IPW (adherence adjusted) estimators were designed for as-treated analyses (Young et al., 2010; Xiao et al., 2010; Westreich et al., 2012; Young and Tchetgen Tchetgen, 2014). Young and colleagues (Young et al., 2019) recently published a general simulation design suitable for assessing IPW per-protocol estimators for a trial. This design allowed subjects from both groups to deviate from the protocol, when the true underlying effect is null, but confounded at various degrees. This study primarily focused on assessing the impact of different measurement schedules of the time-varying covariates on the estimation of per-protocol effects (in terms of bias) based on large, simulated data ($N = 200,000$). The few other simulation designs proposed for the IPW estimator of the per-protocol effect were designed for specific scenarios. For example, multiple studies were based on a simulation design mimicking oncology trials, where only the control group subjects were allowed to switch to the treatment group upon disease progression (Morden et al., 2011; Latimer et al., 2017, 2018). In these simulation studies, estimates from IPW methods were found to be highly sensitive to changes in the switching proportion compared to the other methods they considered. Such treatment switching can be considered a special case of protocol deviation if the protocol did not specify a provision to switch treatment.

While the asymptotic properties of IPW estimator of the per-protocol has been known (Robins and Finkelstein, 2000), Biostatisticians writing the statistical analysis plan for pragmatic trials with practical sample size would be interested in knowing when this approach is expected to deviate from asymptotic behaviour, and may produce sub-optimal results. To the best of our knowledge, there hasn't been a comprehensive study to assess the finite sample size properties of the IPW estimator

of the per-protocol comparing various performance measures such as, bias, variability, 95% coverage etc. Also, we could not find a study that quantified whether these IPW estimators are sensitive to different amounts of protocol deviation in both arms (see Discussion section for further description of gaps in the literature and novelty of the current study).

Aims

In this work, we primarily aim to assess such finite sample properties for the IPW estimator of the per-protocol effect for a sustained but static treatment strategy over a follow-up on a survival outcome. To achieve that, we used a comprehensive simulation study, considering different roles of (measured or unmeasured) baseline covariates under different data generating process. As a secondary aim, we assess whether a change in the rate of non-adherence impacts this IPW estimator's statistical properties. We additionally performed several confirmatory simulations when the following parameter are varied: treatment effect size, trial size, event rate, effect of time on outcome, and sparse follow-up measurements. Finally, we include a case study from the Lipid Research Clinics (LRC) Coronary Primary Prevention Trial (CPPT). This trial had a sample size of 3, 550, with nearly 80% of the subjects deviated from the randomization treatment group by the end of the follow-up. The case study assessed whether a long-term reduction of serum cholesterol in hypercholesterolemic men, initially free of coronary heart disease (CHD), would lead to a lower incidence of CHD. This case study demonstrates real-world applications of the estimators under consideration to account for non-adherence.

2 Methods

2.1 Data generation mechanism

Young et al. proposed a simulation mechanism to mimic the structure of a two-armed randomized pragmatic trial comparing two sustained treatment strategies on a survival outcome (Young et al., 2019). In the current work, we simulated data based on our extension of that algorithm that allowed setting a non-null treatment effect in the outcome event generation model (described below). Baseline observations were generated from random draws from assumed distributions with specified parameters, and subsequent observations were realized from specified models that depended on past observations.

This data generating mechanism consists of monthly measurements, indexed by time as $t = 0, 1, \dots, K$. The month $t = 0$ is the baseline visit, with months $t = 1, 2, \dots, K$ corresponding to follow-up visits, $K = 60$ being the end of follow-up. Participants are followed-up with via monthly visits until the outcome event is observed or the end of the follow-up period (baseline and K visits recorded).

Randomization group or treatment strategies Z : At the baseline visit, participants are randomly assigned to receive either the newer treatment ($Z = 1$) or the standard of care treatment ($Z = 0$). Randomization variable Z is randomly generated as a single Bernoulli trial with the mean 0.5 which indicates, on average, half the participants are randomized to the treatment arm.

Baseline covariate or prognostic factor B_i : At the baseline visit ($t = 0$), a vector of baseline covariates B_i are measured for subject i . Measurement of baseline covariates are common in clinical trials and may include demographic characteristics, e.g., age, sex, race, ethnicity, and baseline disease characteristics, e.g., symptom scores, disease duration. For our simulation, one baseline covariate, B_i for subject i , is assumed to follow a beta distribution: $B_i \sim \text{beta}(2, 2)$. We also considered situations where this baseline covariate is unmeasured.

Post-randomization covariates: For sustained treatment strategies, postbaseline prognostic factors may affect treatment adherence, and hence time-varying variables $L_{t,i}$ for subject i (e.g., development of comorbid conditions, concomitant therapies, lab tests) are important to measure. At each monthly visit t , participant i is measured for post-randomization covariates $L_{t,i} = (L_{1,t,i}, L_{2,t,i})$.

Time-varying continuous covariate $L_{1,t,i}$ is generated initially from a normal distribution with parameters that depend on an individual's values for B_i , and at time t for individual i , based on previous adherence and covariate histories, that are affecting the current values of $L_{1,t,i}$ as follows

$$\begin{aligned} L_{1,t,i} = & \beta_{10} + \beta_{11}B_i + \beta_{12}A_{t-1,i} + \beta_{13}\text{cumsum}(A_{t-2,i}) \\ & + \beta_{14}\text{cumavg}(L_{1,t-1,i}) + \beta_{15}L_{2,t-1,i} + \beta_{16}t + \epsilon_i, \end{aligned} \quad (2.1)$$

with $\epsilon_i \sim N(0, \sigma)$, where $\text{cumavg}()$ denotes the cumulative average function, and $\text{cumsum}()$ denotes the cumulative sum function. $A_{t-1,i}$ and $A_{t-2,i}$ are the indicators for treatment received for time $t - 1$ and $t - 2$ respectively for subject i ($A_{t,i}$ is defined below in more details).

Time-varying binary covariate $L_{2,t,i}$ is generated from a Bernoulli distribution with mean p_{L2} defined as a function of previous adherence and covariate histories:

$$\begin{aligned} \text{logit}(p_{L2}) = & \beta_{20} + \beta_{21}B_i + \beta_{22}\text{cumavg}(L_{1,t,i}) + \\ & \beta_{23}L_{2,t-1,i} + \beta_{24}A_{t-1,i} + \beta_{25}\text{cumavg}(A_{t-2,i}) + \beta_{26}t. \end{aligned} \quad (2.2)$$

Treatment received $A_{t,i}$: Binary indicator for treatment received (treatment vs. placebo or standard care) for subject i at time t is denoted by $A_{t,i}$. $A_{0,i}$ and $A_{t,i}$ represent the treatment received for i -th subject at baseline and follow-up time t respectively. Participants are considered adherent in month t when treatment in that month is the same as the treatment assigned at randomization (Z). When a subject's treatment in month t is not the same as the treatment assigned at randomization for the first time, we label that person as non-adherent at the month (t) due to protocol deviation. We assumed that once a participant becomes non-adherent (e.g., deviates from protocol), they remain non-adherent. $A_{t,i}$ was generated from a Bernoulli distribution with mean p_{Ati} defined by arm. For the standard of care arm:

$$\text{Arm } Z = 0 : \quad \text{logit}(p_{Ati}) = \alpha_{00} + \alpha_{10}\text{cumavg}(L_{1,t,i}) + \alpha_{20}L_{2,t-1,i} + \alpha_{30}t_i + \alpha_{40}B_i. \quad (2.3)$$

Similarly, for the new treatment arm:

$$\text{Arm } Z = 1 : \quad \text{logit}(p_{Ati}) = \alpha_{01} + \alpha_{11}\text{cumavg}(L_{1,t,i}) + \alpha_{21}L_{2,t-1,i} + \alpha_{31}t_i + \alpha_{41}B_i. \quad (2.4)$$

In the above two equations, α_{00} and α_{01} are the intercept parameters from two treatment arms. These parameter values can be altered to control the cumulative proportion of protocol deviations

by the end of the follow-up. In practical scenarios, expectation of perfect adherence may not be realistic, and therefore, trialists often adopt the definition of ‘satisfactory adherence’ (e.g., taking 80% of the medication in a given month) (Hernán and Robins, 2017).

Outcome: We consider a survival outcome (time to an event, e.g., death), where each subject is followed up until an event occurs or the end of 60-th month (study end). Let $Y_{t+1,i}$ be the indicator of the event at month $t+1$ for subject i , where this outcome was generated from a Bernoulli distribution with mean $p_{Y_{ti}}$. In the simulation, if the outcome event is realized (e.g., when $Y_{ti} = 1$), we discontinue data generation for the subject in the following time periods. We assume no loss to follow-up and all participants are at risk of the outcome event at the baseline visit $Y_0 = 0$.

$$\text{logit}(p_{Y_{ti}}) = \theta_0 + \theta_1 B_i + \theta_2 A_{ti}, \quad (2.5)$$

where θ_0 determines the baseline event rate, and θ_2 is the treatment effect of interest (allows non-null effect), while θ_1 is the coefficient associated with the baseline confounder (measured or unmeasured, depending on the scenario considered).

Estimators under consideration

This work considers naive estimates (ITT and per-protocol), conditional per-protocol estimates, and IPW per protocol estimates. For simplicity, we have dropped the subscript i from the following discussions.

Intention to Treat (ITT): The ITT estimate compares all patients who were randomized to the treatment arm vs. all patients who were randomized to the control arm. To calculate model-based ITT estimates, we implemented pooled logistic regression as follows:

$$\text{logit}(P(Y_t = 1)) = \gamma_0 + \gamma_Z Z + \gamma_t t. \quad (2.6)$$

The resulting estimated coefficient γ_Z is the log(OR), estimating ITT effect. Pooled logistic regression with time as a covariate is a suitable outcome model in this situation as pooled logistic regression approximates Cox regression when the outcome of interest is rare (Green and Symons, 1983; D’Agostino et al., 1990; Ngwa et al., 2016) (we have later checked the impact on the estimates when the assumption is not met). Including time as a covariate in pooled logistic regression allows the effect of time on the outcome to be taken into account (D’Agostino et al., 1990; Ngwa et al., 2016).

Naive Per-protocol (Naive PP): We have calculated a per-protocol estimate by censoring participants at the time they become non-adherent to their randomized treatment. The naive model-based per-protocol estimates are calculated using the following pooled logistic regression on the trial data that has been censored based on adherence:

$$\text{logit}(P(Y_t = 1)) = \gamma_0 + \gamma_{A_{pp}} A + \gamma_t t. \quad (2.7)$$

The estimated coefficient $\gamma_{A_{pp}}$ is the log(OR), estimating naive per-protocol effect.

Inverse Probability of (Adherence) Weighted Per-protocol (IPW PP): Inverse Probability Weights adjust for the bias introduced by artificially censoring the data using both the post-randomization covariates L_1 and L_2 and the baseline covariate B . The stabilized weight (sIPW) are calculated as:

$$W_{j,t} = \prod_{j=1}^t \frac{P(A_t = 1|B, t)}{P(A_t = 1|B, t, \bar{L}_{1,t}, \bar{L}_{2,t})}, \quad (2.8)$$

whereas the unstabilized weight (IPW) formula's numerator in equation (2.8) is replaced by 1. The numerator probabilities for stabilized weight formula are the predicted values from logistic regression of the following form:

$$\text{logit}(P(A_t = 1)) = \nu_0 + \nu_t t + \nu_B B, \quad (2.9)$$

and the denominator probabilities for both stabilized and unstabilized formulas are the predicted values from logistic regression of the following form:

$$\text{logit}(P(A_t = 1)) = \nu'_0 + \nu'_t t + \nu'_B B + \nu'_{L_1} \bar{L}_{1,t} + \nu'_{L_2} \bar{L}_{2,t}. \quad (2.10)$$

These weights can then be used to calculate the final IPW per-protocol estimate. The artificially censored trial data is used in the following weighted outcome model:

$$\text{logit}(P(Y = 1)) = \gamma_0 + \gamma_{A_{ipw}} A + \gamma_t t + \gamma_B B. \quad (2.11)$$

Logistic regression was used for calculating the stabilized weights, separately for each trial arm. To estimate stabilized inverse probability weighted per-protocol (sIPW PP), the outcome model (weighted pooled logistic regression) must also include the baseline covariate B . For unstabilized weights (uIPW PP), we should not include B in the outcome model. It is well-known that the estimates using the stabilized weights are associated with lower variance of the effect estimates (Cole and Hernán, 2008), and hence more popularly used. The estimated coefficient $\gamma_{A_{ipw}}$ is the log(OR). When B was unmeasured, the term B was omitted from both weight and outcome model calculations (from equations (2.9) - (2.11)).

Conditional Per-protocol (Adj. PP): We have included a number of conditional per-protocol estimators, where naive per-protocol data (created based on censoring of subjects at the time of protocol deviation) are adjusted by following list of covariates in the pooled logistic regression: (i) baseline adjusted-per-protocol estimate (B Adj. PP; naive PP with adjusting for B in the outcome model), (ii) post-baseline prognostic factor adjusted-per-protocol estimate (L Adj. PP; naive PP with adjusting for L in the outcome mode), and (iii) adjusted by both (B+L Adj. PP; naive PP with adjusting for B and L in the outcome mode) in the outcome model. None of these models were weighted. Appendix G describes rationale of adjusting for the baseline covariates in a pragmatic trial context.

Target parameter for simulations, and estimation method: For all simulations, the treatment effect is quantified using the logarithmic transformation of Odds Ratios (log(OR)s). The pooled logistic regressions are giving us ORs that are approximating hazard ratios under rare event assumption. Standard errors for these estimates were sandwich estimators that take into account the clustering of multiple observations from the same individual (Cameron and Miller, 2015).

2.2 Simulation settings considered

We considered simulations under two broad categories, as per our aim in this manuscript. For our base scenario (Diagram 1 (i); the most complicated case), we set $n = 1,000$ for each arm. Complete set of parameters for the base case is described in the footnote of Figure 1. Appendix Table C.1 describes the set of parameters or characteristics changed to generate each simulation setting. 1,000 iterations were considered in each simulation.

Aim 1: Assessing the impact of different structure of the data generating mechanism: The corresponding relationships between each quantity in our data generating process are summarized in Figure 1. Diagram 1 (i) is our base data generating mechanism, which represents the most complicated scenario, where B is a baseline prognostic factor, directly affecting adherence to the protocol A , post-baseline risk factors L and survival. In Diagram 1 (i), adherence to protocol A affects subsequent post-baseline risk factors L , whereas in Diagram 1 (ii), there is no such effect; and this distinction is true for all the diagrams labeled (i) vs. (ii). Diagrams 2 (i) and (ii) are similar to Diagrams (i) and (ii) respectively, but B does not directly impact adherence to the protocol A but affects post-baseline risk factors L , that could in turn affect adherence A . In Diagrams 3 (i) and (ii), B is a variable that only impacts survival, whereas, B is a noise variable (has no impact on any of the variables) in Diagrams 4 (i) and (ii). In all scenarios, we chose the parameter values such that the cumulative proportion of protocol deviations are approximately the same in both arms (e.g., approximately 40% in both arms; controlled by α_{00} and α_{01}), and approximately the same baseline event rate (e.g., approximately 15% per scenario), determined by θ_0 . Fixing these 3 parameters (via grid search) allowed us to control for variability caused by these extraneous reasons.

Aim 2: Assessing the impact of non-adherence: We considered different non-adherent rates (between 0% to 80%) for the diagrams when adherence affects future time-dependent (all Diagrams with (i) setting). In our simulation, we changed non-adherence rates by the end of the follow-up by changing the intercepts of the parameters α_{00} and α_{01} . We considered three scenarios: (a) varying non-adherence rates when the non-adherence rates are the same in both arms, (b) allowing differential non-adherence rates in both arms, (c) considering sparse measurements (varying degree of gaps in months between each measurement) in recording the following variables: (i) adherence measurements, (ii) adherence and post-baseline prognostic factors, and (iii) post-baseline prognostic factors. In our setting, sparse measurements for the time-varying covariates and adherence were imputed using Last Observation Carried Forward (LOCF) using the most recent prior month's observed values. After imputing, the estimation models were fitted.

Additional simulations to study finite sample properties: Besides these simulations that were assessing our above 2 aims, we also have conducted a number of additional simulations based on our base case (Diagram 1 (i)) for checking the validity of our simulation algorithm, and to assess whether we get expected results in the scenarios when the following parameter are varied:

(1) treatment effects and trial size, (2) event rate, (3) effect of time on outcome, and (4) sparse follow-up measurements. While most of these settings are confirmatory in nature, we are particularly interested about the event rate setting, as the pooled logistic regressions were used under the rare event assumption (approximating Cox regression), and we want to assess the impact on the estimates when the event is not rare.

2.3 Measures of performance

For each effect estimate under consideration, the following statistical properties have been calculated and compared: bias, coverage probability, mean squared error, convergence, bias-adjusted coverage probability, empirical standard error, average model standard error, power or type I error, and confidence interval length (Morris et al., 2019). See Appendix A for definitions and formulas for these performance measures.

3 Results

3.1 Aim 1: Different structures of data generating mechanism

Impact on bias

For the six data generating mechanisms we have considered, we reported bias for each scenario in Appendix Figure D.1. We also showed the impact on bias when baseline variable B is either measured or unmeasured.

Naive estimates: The panels on the left are reporting ITT and naive per-protocol estimates that do not take any covariates (baseline risk factors, or post-baseline prognostic factors) into consideration. For ITT, the treatment effect estimate is unbiased, if there is no treatment effect ($\theta_2 = 0$). This result is consistent with previously reported results (Young et al., 2019). However, for any non-null effect, as expected, we observe bias in the ITT estimates.

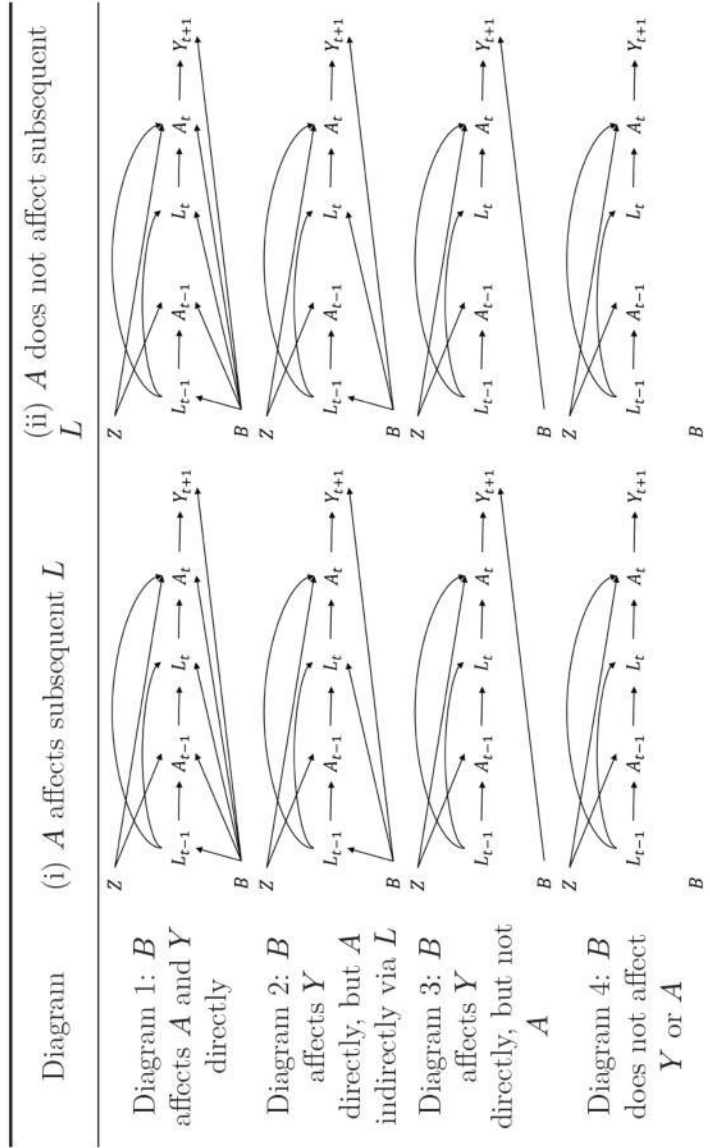


Figure 1: Data generating mechanisms describing the desired simulations for comparing sustained treatment strategies. Z is an indicator of randomization group and B is the baseline covariate. L_{t-1} and L_t are time-varying covariates at times $t-1$ and t , respectively, A_{t-1} and A_t indicates treatment received at times $t-1$ and t , respectively. Y_{t+1} is the outcome of interest observed at time $t+1$. We assumed no loss to follow-up, and once a subject deviates from protocol, they remain non-adherent for the remaining follow-up.

In contrast, the bias in naive per-protocol estimate is considerably higher when B is acting as a confounder (where B directly affects A in Diagrams 1 (i), 1 (ii)). When B is indirectly affecting A (via L) in Diagrams 2 (i) and 2 (ii), the amount of bias is slightly lower than the previous scenarios, but still substantial bias remains. The amount of bias is drastically reduced when B is merely acting as a risk factor for the survival (Diagrams 3 (i) and (ii)), but the bias in the estimation persists when we are dealing with non-null per-protocol effects. Part of this remaining bias is due to not adjusting for artificial censoring in the naive estimation procedure, where the survival could be explained by B . The only situations where naive per-protocol methods provide unbiased estimates is when B is a noise variable (Diagrams 4 (i) and (ii)).

Per-protocol estimates adjusting for covariates: For Diagrams 1 and 2, by the back-door path criterion, baseline prognostic factor B is inducing confounding, and unadjusted treatment effect estimates for the per-protocol will be confounded. Adjustment of B is necessary to produce unbiased treatment effect estimates. Conditioning or IPW could both be used for adjusting B . We obtained three regression estimates, adjusting for B only, L only, or B and L . Particularly when adherence affects future time-dependent variables and the relationship between adherence and survival is confounded by B (Diagrams 1 (i) and 2 (i)), the bias is very high if B is not adjusted. As long as B is adjusted, in combination with L or not, we can obtain unbiased estimates. We can also adjust via inverse probability weighting to obtain unbiased estimates.

Per-protocol estimates when baseline prognostic factors are unmeasured: We also reported the estimates of the per-protocol effects, but in a situation when B was not measured and can not be adjusted. Only if B is a noise variable, adjustment of this variable is not necessary. When adherence does not affect future time-dependent variables or B is merely a risk factor of the survival, the unbiased effect estimates can be estimated only if the treatment effect is null. For non-null effects in these situations, increasing bias levels are observed as the treatment effect increases. When adherence affects future time-dependent variables and B is a confounder, unbiased estimation of the null treatment effect is not guaranteed.

IPW approach performed as well as, or better than the regression adjustment approach (L adjusted) in all scenarios under consideration where B is unmeasured. Compared to conditional regression approach (L Adj. PP), IPW approach is particularly helpful in reducing bias when adherence affects future time-dependent variables and the relationship between adherence and survival is confounded by unmeasured B (Diagram 1 (i) and 2 (i)).

Impact on coverage and mean squared error

When B is measured and adjusted (B Adj. PP and sIPW PP), we observed desirable statistical properties, e.g., nominal coverage and low MSE (Appendix Figure D.2 A). When B is a confounder but unmeasured, adjustment of the baseline variable is not possible. In that case, naive per-protocol estimates provide very low coverage probability as well as high MSE. When only time-varying prognostic factors are adjusted (L Adj. PP), for Diagrams 1 (i) and 2 (i), the coverage probabilities are very low. When we estimate stabilized IPW (in absence of B), the estimates for the coverage and MSE are less than ideal for the data generating process when B is a confounder (Diagrams 1 (i) and 1 (ii) for all parameters, and Diagrams 2 (i) and 2 (ii) for non-null parameters). However,

based on Appendix Figure D.2 B, the worst-case scenarios (i.e., most extreme deviations) of these IPW estimates compared to an ideal scenario (Appendix Figure D.2 A) are much better than those worst-case scenarios observed in the other per-protocol estimates (e.g., L Adj. PP) in terms of MSE. However, the IPW estimates for Diagrams 1 (i) and (ii) were associated with notable under-coverage, and higher standard error.

3.2 Aim 2: Impact of varying non-adherence rates

Varying non-adherence rates

Appendix Figure D.3 described the situation when we change the intercepts of the equations (2.3) and (2.4) to obtain different non-adherence rates by the end of the follow-up, but the rates of non-adherence remain the same in both arms. We only show results from the data generating process when adherence affects future time-dependent variables (all Diagrams with (i) setting). For most of the per-protocol estimates (estimates conditional on B or B and L , and stabilized IPW), the estimates are unbiased even at 60% non-adherence rate (Appendix Figure D.3 A). For unstabilized IPW, however, bias is observed at a lower rate of non-adherence (after 40% in our context). For all the estimators, the model SE generally increases with the increase of the non-adherence rates (Appendix Figure D.3 B).

Differential non-adherence rates

Appendix Figure D.4 considers the situation for the data generating process when adherence affects future time-dependent variables, but the rates of non-adherence can be different in both arms. Similar to the previous graph, we can see that the bias level increased when the non-adherence rates in the control arm become too high. With increased non-adherence rates in the treatment arm, bias increases slightly, only noticeable at a very high rate of non-adherence in the control arm. Particularly for unstabilized IPW estimator, the bias level increases more steeply compared to that for unstabilized IPW or baseline adjusted per-protocol estimator. We also considered another set of simulations when the baseline prognostic factor B is unmeasured. Various per protocol estimates are reported in Appendix Figure E.9. These estimates were associated with slightly higher bias in every situation, but the pattern remains very similar.

Varying non-adherence and varying measurement schedule

Appendix Figure D.5 illustrates the bias as the non-adherence rate varies for different numbers of months between measurements. In this simulation, non-adherence rates were set to be equal in both treatment arms, the effect of interest was fixed at $\theta_2 = -0.7$, and LOCF was implemented before the analysis, and the data generation process follows Diagram 1 (i) (the base scenario). Results presented are based on having both the receipt of treatment A and the time-varying covariates L subject to sparse measurement.

The stabilized IPW per-protocol estimate shows bias increases as the non-adherence rates increase. When non-adherence exceeds approximately 40%, we see the bias increase for a measure-

ment schedule of every 24 months. In contrast, more frequent monthly measurement schedules do not see this increase in bias until non-adherence rates exceed 60%. Variability of the bias increases as non-adherence rates increase (figure not shown). This indicates the high rates of non-adherence and infrequent measurements degrades performance faster than high rates of non-adherence on their own. In contrast, the baseline adjusted per-protocol estimate is approximately unbiased for all non-adherence rates, with measurements taken every 24 months yielding slightly higher bias than monthly measurements. Sparse measurement schedules resulted in moderate and mild increases in bias for the stabilized IPW and baseline adjusted per-protocol estimates respectively, only when adherence was one of the variables subject to sparse follow-up. When only the time-varying covariates were subject to sparse follow-up, all estimates had the same performance in terms of bias regardless of measurement schedule for the data generating mechanism we considered.

3.3 Results from simulations to study finite sample properties

Results of these simulation scenarios based on our base case (Diagram (i)) are outlined in Appendix E. These simulation results can be summarized into the following messages.

Treatment effects and trial size: We considered 3 different trial sizes - 200 (small), 1000 (moderate), 2000 (large) participants per treatment arm. As expected, we observed stabilized IPW and baseline adjusted per-protocol estimators provide approximately unbiased results with moderate to large sample sizes, but some observed bias when the trial size is small. In terms of SE, the gain from moderate to large trial size is not much, but the small trials show high SE, resulting in much reduced power for small trial sizes. Interestingly, the coverage remained nominal (near 95%) even when trial size was small for baseline adjusted or IPW PP.

Event rate: For model-based estimates, the survival model is approximated by pooled logistic regression under the rare event assumption. However, when we varied event rates between 1 – 75%, stabilized IPW and baseline adjusted per-protocol estimators provided unbiased results, with reasonable model SE. We expected cumulative survival estimator to perform better under rare event scenario, but surprisingly model-based estimates performed better in terms of bias. This is partly due to the convergence was slightly lower for cumulative survival estimator.

Effect of time on outcome: When we modified the effect of time on the outcome model, resulting in higher event rates, the stabilized IPW and baseline adjusted per-protocol estimators were not affected by this change, and the results were unbiased. Interestingly, naive estimators (ITT and naive PP) were noticeably impacted by this change. Due to the increase in event rates with the increment of the effect of time, model SEs decreased for all estimators.

Sparse follow-up measurements: As seen in the previous section, with LOCF imputation on the unmeasured records, for non-null treatment effect scenarios, the stabilized IPW and baseline adjusted per-protocol estimates indicate an increase in bias trend as the measurements are less frequently measured over the follow-up period. We also considered complete case analysis. As expected, complete case analysis results consistently had less power compared to LOCF imputation results for the per-protocol estimators.

4 Case Study: Lipid Research Clinics Coronary Primary Prevention Trial

Study setting

The Lipid trial was a two-armed double-blind clinical trial, supported by the National Heart, Lung, and Blood Institute. The study aimed to test the hypothesis that long-term reduction of serum cholesterol in hypercholesterolemic men initially free of coronary heart disease (CHD) would lead to a lesser incidence of CHD (Lipid Research Clinics Program, 1984). The study population was 35-59 years aged asymptomatic, nondiabetic, normotensive men with primary Type II hyperlipoproteinemia, assigned to treatment with cholestyramine or a placebo. The study initially recruited 3,806 participants between mid-1973 and mid-1976, a total of 3,550 subjects were deemed eligible and were randomized at the fifth visit and followed a minimum of seven years. Further details can be found elsewhere (The Lipid Research Clinics Program, 1979).

Outcome, adherence and covariates

The outcome of interest was atherosclerotic CHD death or non-fatal myocardial infarction (Lipid Research Clinics Program, 1984). The percentage of participants who experienced the event of interest was 7.3%. Overall, 84.0% participants in the treatment arm were nonadherent by their last observation, while it was 77.2% in the placebo arm. Many baseline and time-varying covariates were collected in the Lipid trial. We considered five baseline and 38 time-varying covariates in the present study as was done in previous studies (The Lipid Research Clinics Program, 1979; Lipid Research Clinics Program, 1984; Wanis et al., 2020). The description of these covariates and the trial can be found in Appendix F.

Results

Using the discrete-time hazards model, approximated by the pooled logistic regression, we reported log-odds ratio (OR), robust SE, OR, and 95% CI from the estimators under consideration in Appendix Table F.1. Our adjusted per-protocol analysis using the stabilized version of the IPW PP shows that the risk of CHD death or non-fatal myocardial infarction in the cholestyramine treatment group was 26% less compared to placebo. Utilizing unstabilized weights (with maximum weight 172) resulted in a somewhat different HR estimate compared to those obtained via baseline risk factor adjusted per-protocol and stabilized inverse probability of adherence weighted per-protocol estimators. In contrast, when we truncated 5% of the weights on both extremes, the maximum weight reduced to 1.44, and the resulting HR (24% reduction in risk) was very similar to the one obtained using the stabilized weights.

5 Discussion

Different data generating mechanisms This work compared various per-protocol estimators under different data generating mechanisms including: whether baseline confounding exists, and whether time-dependent confounding exists where a post-baseline confounder is impacted by prior adherence. Based on our simulations, we observed that no matter whether we are using baseline adjusted per-protocol or stabilized IPW per-protocol estimators, we end up getting the correct answer if all the necessary baseline confounders are measured, adjusted and model is correctly specified and time-varying confounders are measured frequently during the follow-up period. And these findings are supported by the previous literature (Robins and Finkelstein, 2000; Young et al., 2019). In practice, however, it is challenging to guarantee measurement and adjustment of all necessary confounders. Our simulation assessed realistic conditions where some confounders might not be measured or adjusted (i.e., a realistic situation that has not been explored earlier). Based on the data generating mechanisms under consideration, we observed that even the worst-case scenario for the IPW per-protocol estimate (in terms of performance measures: bias, MSE and coverage) is better than the conditional estimators, when unmeasured confounding exists. In other words, in the complex realistic situations, even if the analysts do not have access to all confounders or may not know the true data generating mechanism, IPW per-protocol is the preferable estimator for the practitioners in most situations. Subject area experts may have some idea about a possible data generating mechanism by which the variables under consideration impact each other, variables that are relevant for adjustment, and whether those variables are measurable. Our results highlight potential limitations of adjusted pre-protocol estimates in terms of the finite sample properties.

Different adherence patterns In our simulation, when all baseline confounders are measured, and adherence patterns are the same in both arms, both baseline adjusted per-protocol and IPW per-protocol estimates obtained negligible bias as non-adherence rates increased up to 60%. However, for even higher non-adherence rates, the degree of bias depends on the data generating mechanism. When a baseline confounder impacts not only adherence, but also time-dependent confounders, the degree of bias is the highest. When the baseline confounder only impacts time-dependent confounders, the level of bias is slightly reduced compared to the previous scenario. When the baseline covariate is not a confounder, the level of bias is minimal. These results align with previous work while also highlighting that different underlying data generation mechanisms greatly impact the performance of adjusted per-protocol effect estimates.

When the adherence pattern differs between arms (i.e., another realistic situation that has not been explored earlier), the estimates showed similar trends, but the degree of bias may decrease if the adherence pattern in either arm is low. We have also considered a differential adherence scenario without adjusting for a necessary confounder (e.g., when we have an unmeasured baseline prognostic factor), and the results were slightly more biased, but the trends remained the same. In all these settings, we have seen that the unstabilized version of the IPW estimator for high non-adherence is associated with more bias when baseline confounder impacts not only adherence, but also time-dependent confounders. However, when we estimated the stabilized version of the IPW estimators, the level of bias reduced drastically. With sparse follow-up measures, even the stabilized

IPW per-protocol estimates saw bias increase for lower non-adherence rates than with more frequent measurement of the post-baseline variables, due to their reliance on the measurement and adjustment of the post-baseline prognostic factors. When a trial is subject to both high rates of non-adherence and sparse measurements during the follow-up period, and all the necessary baseline confounders are measured, baseline adjusted per-protocol estimates were associated with the least bias. As the true data generating process is generally unknown and guaranteeing measurement of all the necessary confounding is rarely possible, estimating different per-protocol estimators may provide us useful insights.

Confirmatory simulations Our confirmatory simulations used the base case where we have a baseline confounder, and there exist post-baseline confounders that are affected by prior treatment. Using this simulation mechanism, we have the following observations. (1) Trials with a low sample size (e.g., $n = 200$ participants per arm according to our simulations) often produce unbiased adjusted per-protocol effect estimates with sufficient coverage, but to achieve sufficient power, larger trial sizes are required. (2) In trials where the event rate is low (e.g., $< 1\%$ according to our simulations) adjusted per-protocol estimates are biased with high variability. Model-based adjusted per-protocol estimates did not have any issue with convergence with low event rates, while cumulative survival type estimates have lower variability and convergence. We have utilized pooled logistic regression models to approximate survival models with the “rare disease assumption” (Murray et al., 2021). Interestingly, in simulations with large event rates, the results from model-based estimates were unbiased, and this illustrates that the results were not too sensitive to violation of this rare disease assumption in the settings we have considered. (3) Differing effects of time on the outcome did not impact the IPW per-protocol estimates. (4) By comparing different measurement schedules during the follow-up period, we have shown that the stabilized IPW and baseline adjusted per-protocol estimates show a slight increase in bias as the frequency of measurements decreases in the follow-up period. This observation indicates the importance of frequent measures of both adherence and time-varying covariates to facilitate the calculation of IPW per-protocol estimates. A similar conclusion was drawn previously for the null treatment scenario (Young et al., 2019), and our conclusion is extended for a generalized non-null scenario. We have also compared our estimators’ performance when there are sparse follow-up measurements, which are imputed using LOCF or analyzed as a complete case. Our work illustrated that LOCF estimates have lower variability and bias than complete case estimates as the measurement schedule becomes more infrequent. LOCF imputation also yields greater power than complete case as the treatment effect varies. Similar conclusions were previously drawn for as-treated analysis (Mojaverian et al., 2015), and we extended this observation in the per-protocol analysis setting.

Novelty of the current work We extended our simulations for non-null situations compared to the previous simulation by Young et al. (2019) focusing on bias assessment of the per-protocol effect estimators (Young et al., 2019). The previous study generated one large cohort (with a sample size of 200,000) to focus on bias (Young et al., 2019). In this work, each simulated trial had 2,000 in most setting, allowing us to investigate variability and coverage of the effect estimates as measures

of performance.

The current work is the a comprehensive simulation study considering many settings, and first one to consider many data generating mechanisms (e.g., how a potential prognostic factor measured baseline impacts other variables) and different adjustment strategies (e.g., when that baseline variable is available to adjust or not) that were not considered in prior studies in understanding the finite sample properties of the estimators under consideration.

In terms of estimation methods, previous works were comparing the performances of cumulative survival effect estimates (Murray and Hernán, 2018; Young et al., 2019). One disadvantage of the cumulative survival version of the per-protocol estimator is that it has no direct mechanism to include the baseline covariate, and hence utilizing stabilized version of the weights are not possible in a straight-forward manner. Model-based estimators are free from such limitation. Our work primarily focuses on the performance of model-based effect estimates, although cumulative survival effect estimates were also calculated for confirmation purposes (see Appendix B for details). Even though estimates from pooled logistic regression is typically described as approximating estimates from a survival analysis model when the event rate is rare, this work showed unbiased model-based estimates resulted in scenarios with a wide range of event rates. This statement remained true when event rates were not rare. Hence we feel comfortable that the results of the simulation are generalize to situation when the event rate is not rare.

In previous simulations, it was reported that the IPW per-protocol estimator is sensitive to a high proportion of control patients (more than 65%) switching to treatment group (Latimer et al., 2017), but produce low bias when switching proportions are low (less than 40%) (Latimer et al., 2018). The current work is the first in our knowledge to assess if similar patterns occur in a general scenario, where “both treatment groups” are allowed to deviate from the protocol, and if such patterns exist, under what condition the bias is most severe. We additionally considered the deferential non-adherence setting which is very common in realistic settings. These settings included a situation with high non-adherence rate (e.g., 80%) that was aligned with the non-adherence rate observed in our case study.

Case study Using the intention to treat analysis, a 15% reduction in risk of CHD death or non-fatal myocardial infarction was previously reported in the cholestyramine treatment group than placebo (Lipid Research Clinics Program, 1984). This result is consistent with our ITT result. Naive per-protocol estimate showed a 20% reduction in risk. However, one could argue that this result is subject to artificial (i.e., dependent) censoring as the analysis required censoring the patients at the time of medication non-adherence for the first time, without adjusting for the prognostic factors that could be predictive of medication non-adherence. When we adjusted for the post-baseline prognostic factors in a conditional model, the results deviated substantially. HR moved to the other direction of the null (20% increase in risk), but the 95% confidence interval was much wider. It is possible that there exist some time-dependent confounders that is affected by past treatment, like our Diagrams with the (i) settings, and hence, adjusting for these post-baseline factors in a conditional model resulted in very different estimates. Estimates obtained by baseline adjusted approaches (baseline risk factor adjusted per-protocol and stabilized inverse probability of adherence weighted

per-protocol) resulted in similar conclusions (22% and 26% reduction in risk).

As in any real data analysis, despite the use of many baseline and post-baseline covariates, we can not still guarantee that no open back-door paths exist. However, the similarity of several different estimates (from baseline risk factor adjusted per-protocol, truncated unstabilized and stabilized inverse probability of adherence weighted per-protocol) is comforting in this case study.

Strengths A strength of this work is the number of different practical scenarios (e.g., different data generating mechanisms, different adherence patterns) to which these effect estimates have been applied, and consideration of different data generating mechanisms. We primarily focused on the model-based effect estimates (as opposed to cumulative survival effect estimates). These estimators allow the direct estimation of the variance of the treatment effect and the easier inclusion of any potential baseline covariates. These features of the model-based approaches allowed us to report additional statistical properties for each effect estimate including power, average model standard error, empirical standard error, coverage probability, and unbiased coverage probability. These statistical properties provide strong evidence of the performance of each effect estimate beyond bias. We also assessed the finite sample properties with practical sample sizes. Even though there exist large randomized pragmatic trials (e.g., Cocoros et al. (2018) had 44,786 subjects), statistical properties of relatively moderately size pragmatic trials are of much interest.

Limitations and assumptions Similar to the previous simulation, we assume that no participants were lost to follow-up and that once a participant became non-adherent, they remained non-adherent for the remainder of the trial (Young et al., 2019). For events such as death, it is usually possible to get a reliable estimate of death times from vital statistics registries. However, there might be situations when censoring may be appropriate to account for loss to follow-up. These assumptions may limit the generalizability of our findings. Estimators under consideration assume that all important prognostic factors are measurable, but in real-world studies it may not be guaranteed. If important prognostic factors that affect adherence or loss to follow-up are not measured and adjusted, the obtained effect estimates may be biased (Hernán and Robins, 2017). We did however, assess the performances of the estimators when such a factor remains unmeasured. We also considered a limited number of covariates and assumed correct specification of the models in the data generation and estimation.

In our analyses, we used LOCF for imputation. Despite the popularity of the LOCF, compared to more modern imputation approaches such as multiple imputation or mixed-effect model repeated measure imputation, LOCF may produce biased results (Siddiqui et al., 2009; Mojaverian et al., 2015; Moodie et al., 2008). Recent research suggested that analysts should consider multiple imputation approach for per-protocol analyses in pragmatic trials in a sparse follow-up situation (when missing completely at random is plausible) (Karim and Hossain, 2022). Particularly the multiple imputation approach is necessary when confounding exists and higher variability of the time-varying factor is evident. Future research could consider the performance of these estimators under the missing at random assumption. Some of the future works are proposed in Appendix H.

Conclusion In the current study, we explored the relative performance of different per-protocol effect estimators in the presence of non-adherence under different data generation mechanisms for sustained treatment strategies. These mechanisms included different roles of a baseline variable, and future time-varying prognostic factors potentially being impacted by past adherence. We also compared performance when the necessary baseline variable is measured or not. In all settings, when baseline confounders are measured and adjusted, we generally obtain unbiased estimates when IPW per-protocol or the baseline adjusted per-protocol estimators are used. However, when some of these variables are not measured, IPW per-protocol may still be preferable compared to the other estimators under consideration. We also observed that high non-adherence patterns might impact these effect estimators, particularly when the data generating process includes complex patterns (existence of baseline or time-dependent confounding that may be impacted by previous adherence history).

Acknowledgements

We thank Lang Wu (Department of Statistics, The University of British Columbia) and Hubert Wong (School of Population and Public Health, The University of British Columbia) for helpful comments. This research was enabled in part by computing support provided by WestGrid (www.westgrid.ca), Compute Canada (www.computecanada.ca), and the Centre for Health Evaluation and Outcome Sciences. We also thank Sharon Roman, a patient partner of the BC SUPPORT Unit methods clusters, for her involvement with the project.

Funding details: This work was supported by BC Support Unit's Real-World Clinical Trials Methods Cluster, Project #2, led by Dr. Karim (with research members Paul Gustafson, Joan Hu, Hubert Wong, and Derek Ouyang), and Dr. Karim's Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery Accelerator Supplements.

Disclosure statement: LM and BH declare no potential conflict of interests. MEK is supported by the Michael Smith Foundation for Health Research Scholar award. Over the past three years, MEK has received consulting fees from Biogen Inc. for consulting unrelated to this current work.

Data availability statement The trial dataset access can be requested from National Heart, Blood, and Lung Institute. The analysis of secondary and de-identified data was exempt from the requirements for research ethics approval both in accordance with the University of British Columbia Policy 89 and in accordance with the provisions of the Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans, Article 2.5. Sample software codes used are available from the following website: <https://ehsanx.github.io/IPAW-Per-protocol-Estimator/>.

References

- Cameron, A. C. and Miller, D. L. (2015), "A Practitioner's Guide to Cluster-Robust Inference," *Journal of Human Resources*, 50, 317–372.
- Cocoros, N. M., Pokorney, S. D., Haynes, K., Garcia, C., Al-Khalidi, H. R., Al-Khatib, S. M., Archdeacon, P., Goldsack, J. C., Harkins, T., Lin, N. D., and et al. (2018), "FDA-Catalyst-Using

- FDA's Sentinel Initiative for large-scale pragmatic randomized trials: Approach and lessons learned during the planning phase of the first trial," *Clinical Trials*, 16, 90–97.
- Cole, S. R. and Hernán, M. A. (2008), "Constructing Inverse Probability Weights for Marginal Structural Models," *American Journal of Epidemiology*, 168, 656–664.
- D'Agostino, R. B., Lee, M.-L., Belanger, A. J., Cupples, L. A., Anderson, K., and Kannel, W. B. (1990), "Relation of pooled logistic regression to time dependent Cox regression analysis: The Framingham heart study," *Statistics in Medicine*, 9, 1501–1515.
- Danaei, G., Rodríguez, L. A. G., Cantero, O. F., Logan, R., and Hernán, M. A. (2013), "Observational data for comparative effectiveness research: an emulation of randomised trials of statins and primary prevention of coronary heart disease," *Statistical Methods in Medical Research*, 22, 70–96.
- Gamerman, V., Cai, T., and Elsässer, A. (2019), "Pragmatic randomized clinical trials: best practices and statistical guidance," *Health Services and Outcomes Research Methodology*, 19, 23–35.
- Green, M. S. and Symons, M. J. (1983), "A comparison of the logistic risk function and the proportional hazards model in prospective epidemiologic studies," *Journal of Chronic Diseases*, 36, 715–723.
- Group, C. D. P. R. (1980), "Influence of Adherence to Treatment and Response of Cholesterol on Mortality in the Coronary Drug Project," *New England Journal of Medicine*, 303, 1038–1041.
- Hernán, M. A. and Hernández-Díaz, S. (2012), "Beyond the intention-to-treat in comparative effectiveness research," *Clinical Trials*, 9, 48–55, PMID: 21948059.
- Hernán, M. A. and Robins, J. M. (2017), "Per-Protocol Analyses of Pragmatic Trials," *New England Journal of Medicine*, 377, 1391–1398.
- Karim, M. E. and Hossain, M. B. (2022), "Considerations for choosing an imputation method for addressing sparse measurement issues dictated by the study design: An illustration from per-protocol analysis in pragmatic trials," *Research Methods in Medicine and Health Sciences*, 3, 74–85.
- Latimer, N. R., Abrams, K., Lambert, P., Crowther, M., Wailoo, A., Morden, J., Akehurst, R., and Campbell, M. (2017), "Adjusting for treatment switching in randomised controlled trials—a simulation study and a simplified two-stage method," *Statistical Methods in Medical Research*, 26, 724–751.
- Latimer, N. R., Abrams, K. R., Lambert, P. C., Morden, J. P., and Crowther, M. J. (2018), "Assessing methods for dealing with treatment switching in clinical trials: a follow-up simulation study," *Statistical Methods in Medical Research*, 27, 765–784.
- Lipid Research Clinics Program (1984), "The lipid research clinics coronary primary prevention trial results I. Reduction in Incidence of Coronary Heart Disease," *JAMA*, 251, 351–364.

- Lu, H., Cole, S. R., Hall, H. I., Schisterman, E. F., Breger, T. L., K Edwards, J., and Westreich, D. (2019), “Generalizing the per-protocol treatment effect: The case of ACTG A5095,” *Clinical Trials*, 16, 52–62.
- Mo, Y., Lim, C., Watson, J. A., White, N. J., and Cooper, B. S. (2020), “Non-adherence in non-inferiority trials: pitfalls and recommendations,” *BMJ*, 370.
- Mojaverian, N., Moodie, E. E., Bliu, A., and Klein, M. B. (2015), “The Impact of Sparse Follow-up on Marginal Structural Models for Time-to-Event Data,” *American Journal of Epidemiology*, 182, 1047–1055.
- Moodie, E. E., Delaney, J. A., Lefebvre, G., and Platt, R. W. (2008), “Missing confounding data in marginal structural models: a comparison of inverse probability weighting and multiple imputation,” *International Journal of Biostatistics*, 4, Article 13.
- Morden, J. P., Lambert, P. C., Latimer, N., Abrams, K. R., and Wailoo, A. J. (2011), “Assessing methods for dealing with treatment switching in randomised controlled trials: a simulation study,” *BMC Medical Research Methodology*, 11, 1–20.
- Morris, T. P., White, I. R., and Crowther, M. J. (2019), “Using simulation studies to evaluate statistical methods,” *Statistics in Medicine*, 38, 2074–2102.
- Murray, E. J., Caniglia, E. C., and Petito, L. C. (2021), “Causal survival analysis: A guide to estimating intention-to-treat and per-protocol effects from randomized clinical trials with non-adherence,” *Research Methods in Medicine and Health Sciences*, 2, 39–49.
- Murray, E. J., Claggett, B. L., Granger, B., Solomon, S. D., and Hernán, M. A. (2020), “Adherence-adjustment in placebo-controlled randomized trials: An application to the candesartan in heart failure randomized trial,” *Contemporary Clinical Trials*, 90.
- Murray, E. J. and Hernán, M. A. (2016), “Adherence adjustment in the Coronary Drug Project: A call for better per-protocol effect estimates in randomized trials,” *Clinical Trials*, 13, 372–378.
- (2018), “Improved adherence adjustment in the Coronary Drug Project,” *Trials*, 19.
- Murray, E. J., Swanson, S. A., and Hernán, M. A. (2019), “Guidelines for estimating causal effects in pragmatic randomized trials,” ArXiv:1911.06030.
- Ngwa, J. S., Cabral, H. J., Cheng, D. M., Pencina, M. J., Gagnon, D. R., Lavalley, M. P., and Cupples, L. A. (2016), “A comparison of time dependent Cox regression, pooled logistic regression and cross sectional pooling with simulations and an application to the Framingham Heart Study,” *BMC Medical Research Methodology*, 16.
- Robins, J. M. and Finkelstein, D. M. (2000), “Correcting for noncompliance and dependent censoring in an AIDS clinical trial with inverse probability of censoring weighted (IPCW) log-rank tests,” *Biometrics*, 56, 779–788.

- Schwartz, D. and Lellouch, J. (1967), “Explanatory and pragmatic attitudes in therapeutical trials,” *Journal of Clinical Epidemiology*, 20, 637–648.
- Siddiqui, O., Hung, H. M., and O’Neill, R. (2009), “MMRM vs. LOCF: a comprehensive comparison based on simulation study and 25 NDA datasets,” *Journal of Biopharmaceutical Statistics*, 19, 227–246.
- The Lipid Research Clinics Program (1979), “The coronary primary prevention trial: design and implementation,” *Journal of Chronic Diseases*, 32, 609–631.
- Wanis, K. N., Madenci, A. L., Hernán, M. A., and Murray, E. J. (2020), “Adjusting for adherence in randomized trials when adherence is measured as a continuous variable: An application to the Lipid Research Clinics Coronary Primary Prevention Trial,” *Clinical Trials*, 1740774520920893.
- Westreich, D., Cole, S., Schisterman, E., and Platt, R. (2012), “A simulation study of finite-sample properties of marginal structural Cox proportional hazards models,” *Statistics in Medicine*, 31, 2098–2109.
- Xiao, Y., Abrahamowicz, M., and Moodie, E. (2010), “Accuracy of conventional and marginal structural Cox model estimators: A simulation study,” *The International Journal of Biostatistics*, 6, 1–28.
- Young, J. and Tchetgen Tchetgen, E. (2014), “Simulation from a known Cox MSM using standard parametric models for the g-formula,” *Statistics in Medicine*, 33, 1001–1014.
- Young, J. G., Hernán, M. A., Picciotto, S., and Robins, J. M. (2010), “Relation between three classes of structural models for the effect of a time-varying exposure on survival,” *Lifetime Data Analysis*, 16, 71.
- Young, J. G., Vatsa, R., Murray, E. J., and Hernán, M. A. (2019), “Interval-cohort designs and bias in the estimation of per-protocol effects: a simulation study,” *Trials*, 20.
- Zwarenstein, M., Treweek, S., Gagnier, J. J., Altman, D. G., Tunis, S., Haynes, B., Oxman, A. D., and Moher, D. (2008), “Improving the reporting of pragmatic trials: an extension of the CONSORT statement,” *BMJ*, 337.

Received: August 9, 2022

Accepted: January 8, 2023