Trials

METHODOLOGY

Interval-cohort designs and bias in the estimation of per-protocol effects: a simulation study

Jessica G. Young[1]* , Rajet Vatsa[2], Eleanor J. Murray[3,6] and Miguel A. Hernán[3,4,5]

Abstract

Background: Randomized trials are considered the gold standard for making inferences about the causal effects of treatments. However, when protocol deviations occur, the baseline randomization of the trial is no longer sufficient to ensure unbiased estimation of the per-protocol effect: post-randomization, time-varying confounders must be sufficiently measured and adjusted for in the analysis. Given the historical emphasis on intention-to-treat effects in randomized trials, measurement of post-randomization confounders is typically infrequent. This may induce bias in estimates of the per-protocol effect, even using methods such as inverse probability weighting, which appropriately account for time-varying confounders affected by past treatment.

Methods/design: In order to concretely illustrate the potential magnitude of bias due to infrequent measurement of time-varying covariates, we simulated data from a very large trial with a survival outcome and time-varying confounding affected by past treatment. We generated the data such that the true underlying per-protocol effect is null and under varying degrees of confounding (strong, moderate, weak). In the simulated data, we estimated per-protocol survival curves and associated contrasts using inverse probability weighting under monthly measurement of the time-varying covariates (which constituted complete measurement in our simulation), yearly measurement, as well as 3- and 6-month intervals.

Results: Using inverse probability weighting, we were able to recover the true null under the complete measurement scenario no matter the strength of confounding. Under yearly measurement intervals, the estimate of the per-protocol effect diverged from the null; inverse probability weighted estimates of the per-protocol 5-year risk ratio based on yearly measurement were 1.19, 1.12, and 1.03 under strong, moderate, and weak confounding, respectively. Bias decreased with measurement interval length. Under all scenarios, inverse probability weighted estimators were considerably less biased than a naive estimator that ignored time-varying confounding completely.

Conclusions: Bias that arises from interval measurement designs highlights the need for planning in the design of randomized trials for collection of time-varying covariate data. This may come from more frequent in-person measurement or external sources (e.g., electronic medical record data). Such planning will provide improved estimates of the per-protocol effect through the use of methods that appropriately adjust for time-varying confounders.

Keywords: Per-protocol effect, Inverse probability weighting, Interval cohorts, Simulation

*Correspondence: jyoung@hsph.harvard.edu
[1]Department of Population Medicine, Harvard Medical School & Harvard Pilgrim Health Care Institute, Boston, USA
Full list of author information is available at the end of the article

Background

In randomized trials, the *per-protocol effect* is the effect that would have been estimated if all participants had adhered to their randomly assigned treatment strategies during the entire follow-up [1]. However, because adherence to the assigned treatment strategy is not in itself randomized, a naive comparison that excludes trial participants who fail to adhere to their assigned strategies will generally be biased [2].

For example, in a trial of a new treatment versus standard of care to treat coronary heart disease, adherers to the treatment may be individuals who also tend to take antihypertensive treatment. Thus, a lower rate of disease among adherers may simply reflect their higher uptake of antihypertensives rather than a benefit of the treatment under study. Therefore, analyses that attempt to estimate the per-protocol effect typically need to adjust for prognostic factors that, like antihypertensive use in our example, are also associated with adherence. That is, per-protocol analyses are observational analyses of the randomized trial data and therefore need to adjust for confounders.

In randomized trials of point interventions that are administered shortly after randomization (e.g., a one-dose vaccination, a one-time screening test), adherence to the assigned intervention is fully determined at baseline and therefore can only be affected by baseline factors. The implication is that per-protocol analyses of point interventions only need to adjust for baseline confounders. On the other hand, in randomized trials of treatment strategies that are sustained during the follow-up (e.g., treatment for coronary heart disease, antiretroviral treatment for HIV-positive patients), adherence to the treatment strategy must also be sustained during the follow-up. The implication of this potentially time-varying adherence is that per-protocol analyses of sustained strategies need to adjust for time-varying confounders — time-varying prognostic factors that affect treatment decisions — as well as for baseline confounders — baseline prognostic factors that affect treatment decisions [3–6]. For example, in a randomized trial to estimate the effect of two antiretroviral therapies on mortality, an increased alcohol intake during the follow-up is a time-varying confounder because it affects both the risk of death and of non-adherence to the assigned treatment.

It follows that valid estimation of the per-protocol effect of sustained treatment strategies requires adequate data collection of treatment and confounders after randomization. Many randomized trials collect such post-randomization data, but most only do so at pre-specified intervals (e.g., every 12 months). Because non-adherence may occur at any time during the follow-up, the confounders measured at the pre-specified times may not be sufficient or relevant to adjust for non-adherence that took place at an unknown time between the pre-specified measurement times.

In this paper, we review the impact of interval measurement on the estimation of per-protocol effects in randomized trials [7]. We conduct a simulation study to illustrate the potential magnitude of bias, even using causal inference methods for longitudinal settings such as inverse probability (IP) weighting [8], which appropriately account for time-varying confounders affected by past treatment.

Methods

Simulation design

We simulated data from a hypothetical randomized trial to quantify the effect of a new drug treatment compared to the standard of care on 5-year mortality risk.

Each individual is assigned to either the new drug treatment ($Z = 1$) or to standard of care ($Z = 0$) and followed until death or the administrative end of the study (60 months post-randomization), whichever comes first. We assume the exact month of death is known, as is common when studies link their data with death registries. For simplicity and without loss of generality, no individual is lost to follow-up.

Define $t = 0, \ldots, 60$ as an index of follow-up month with $t = 0$ the month of randomization (baseline). Let Y_t be an indicator of death by month t with $Y_0 \equiv 0$ for all individuals (all participants are alive and therefore at risk of the outcome at baseline) and A_t an indicator of whether the new drug treatment is taken in month t. An individual deviated from the protocol in the first month t in which $A_t \neq Z$. In our simulated study approximately 40% of individuals in both arms deviated from the protocol at some point during the follow-up. Figure 1 shows the cumulative proportion of protocol deviations over the study period by treatment arm.

In randomized trials, treatment A_t will typically depend on both baseline (e.g., sex, race, baseline age) and post-baseline (e.g., lab measurements, concomitant medications) risk factors for the outcome (e.g., death). Let $L_t = (L_{1t}, L_{2t})$ be a vector of such risk factors in month t, with L_{1t} a lab measurement (continuous) and L_{2t} the use of a concomitant medication (binary).

The causal diagram in Fig. 2 outlines the data-generating process of our simulated study. The node U represents a vector of baseline unmeasured outcome risk factors that also may affect L_t (e.g., genetic factors) with no direct effect on treatment at any time (as depicted by the absence of an arrow from U into A_{t-1} or A_t in Fig. 2). As expected in many realistic settings, the time-varying covariates L_t also may be affected by past treatment adherence (as depicted by the arrow from A_{t-1} to L_t in Fig. 2). For example, adherence to the standard versus the new treatment may affect values of future lab measurements.
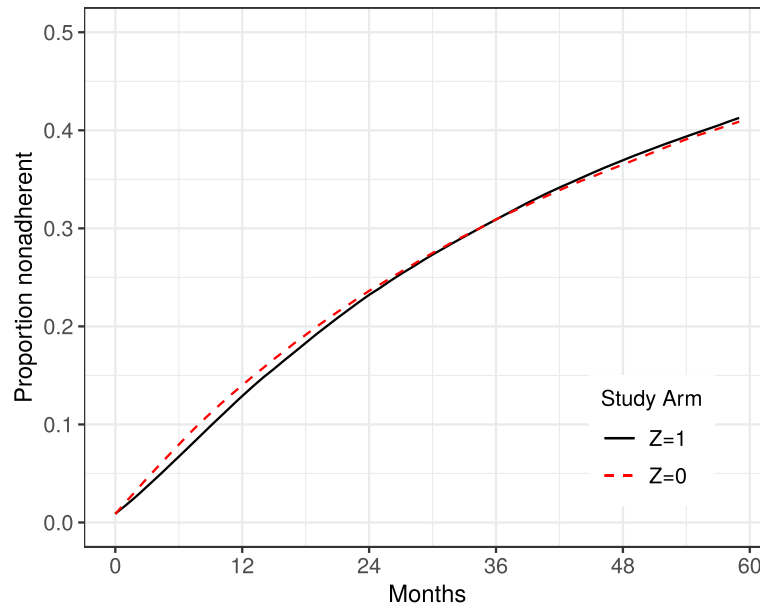
Fig. 1 Proportion of participants who deviate from the protocol over the study period by treatment arm

We generated the data such that 100,000 individuals are assigned to each arm. We quantified bias for a given approach by the difference between the effect estimate obtained by that approach in this very large sample and the true effect value. Had we used a smaller sample size (e.g., 100 individuals assigned to each arm), random variability could explain some differences between effect estimates and the true values of the effect (unless we had used the average over a large number of small samples, which is nearly equivalent to generating a single very large sample — this is illustrated in Additional file 2).

We generated the data such that both the causal effect of treatment A_t for all t and the direct effect of randomization (Z) not mediated through treatment are null, as shown in Fig. 2 by the absence of any *causal paths* (paths consisting of arrows going in the same direction) connecting Z, A_{t-1}, or A_t with the future outcome (Y_{t+1}).

Therefore, both the intention-to-treat effect and the per-protocol effect are null.

Data-generating models

We generated longitudinal data according to the following models for each subject $i = 1, \ldots, 200,000$ ($i = 1, \ldots, 100,000$ assigned $Z_i = 1$ and $i = 100,001 \ldots, 200,000$ assigned $Z_i = 0$): U_i was generated from a uniform distribution between 0 and 1. Then the following were generated for each month $t = 0$ until $t = 59$ or until $Y_{t+1i} = 1$ was generated, whichever came first:

- L_{1ti} was generated from a normal distribution such that $L_{1ti} = 6U_i - A_{t-1i} - \text{cumavg}(\overline{A}_{t-2i}) + 0.25\text{cumavg}(\overline{L}_{1t-1i}) + 0.01t + \epsilon_i$ with $\epsilon_i \sim N(0, \sigma = 2)$, $\text{cumavg}(\overline{A}_{t-2i})$ is the cumulative
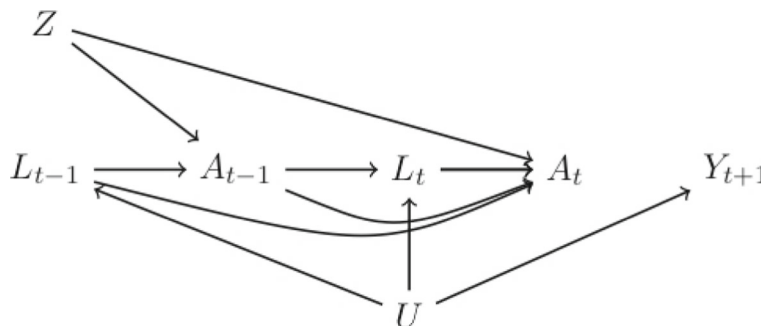

Fig. 2 A causal diagram representing the underlying mechanism for protocol deviations in our study

average of $(A_{0i}, \ldots, A_{t-2i})$, and cumavg$(\overline{L}_{1t-1i})$ is the cumulative average of $(L_{10i}, \ldots, L_{1t-1i})$.

- L_{2ti} was generated from a Bernoulli distribution with mean p_{L2i}, equal to the probability that $L_{2t} = 1$ given individual i's treatment and covariate history and survival to t, defined such that
logit$(p_{L2i}) = -5 + 3U_i + 1.25$cumavg$(\overline{L}_{1ti}) +$
$0.5L_{2t-1i} + 0.25A_{t-1i} + 0.25$cumavg$(A_{t-2i}) + 0.01t$.

- For any individual i deviating from the protocol by $t - 1$ (i.e., $A_{t-1i} \neq Z_i$), we set $A_{ti} = A_{t-1i}$ (once an individual stops complying we assume they stay non-compliant). Alternatively, for any individual i complying with the protocol through $t - 1$ (i.e., all $A_{ji} = Z_i$ for $j < t$), A_{ti} was generated from a Bernoulli distribution with mean p_{Ai}, equal to the probability that $A_t = 1$ given individual i's treatment and covariate history and survival to t, such that

$$\text{logit}(p_{Ai}) = \alpha_0 + 0.4\text{cumavg}(\overline{L}_{1ti}) + 0.35L_{2t-1i}. \quad (1)$$

For individuals assigned $Z_i = 1$ (active treatment), we set $\alpha_0 = 4.0$. For individuals assigned $Z_i = 0$ (standard of care), we set $\alpha_0 = -6.5$.

- The death indicator Y_{t+1i} was generated from a Bernoulli distribution with mean p_{Yi}, equal to the probability that $Y_{t+1} = 1$ given individual i's treatment and covariate history and survival to t, such that

$$\text{logit}(p_{Yi}) = \theta_0 + \theta_1 U_i. \quad (2)$$

We considered three versions of this data-generating mechanism, varying the values of θ_0 and θ_1 in the model (2). As we explain in the section "Defining and estimating the per-protocol effect", given our data-generating models, the magnitude of θ_1 determines the magnitude of time-varying confounding (and θ_0 the baseline event rate). We considered the following variations: "strong confounding" $\theta_1 = 8$ ($\theta_0 = -11$), "moderate confounding" $\theta_1 = 3$ ($\theta_0 = -7$), and "weak confounding" $\theta_1 = 0.5$ ($\theta_0 = -6$). We also considered three variations of the "strong confounding" scenario under different choices of α_0 in model (1) that reduced the chance of deviating from the protocol in both arms. Table 1 displays the cumulative proportion of protocol deviations by the end of the study period by treatment arm resulting from different choices of α_0.

R code implementing this simulation design is provided in Additional file 1.

Defining and estimating the intention-to-treat effect

We can define the intention-to-treat effect for any follow-up month $t + 1 = 1, \ldots, 60$ as a contrast of the cumulative risks in arm $Z = 1$, Pr$[Y_{t+1} = 1|Z = 1]$ versus in arm $Z = 0$, Pr$[Y_{t+1} = 1|Z = 0]$. Our data generation, under

Table 1 Proportion of protocol deviations under different choices of α_0 in (1) by arm under "strong confounding"

Scenario	Arm	α_0	Cumulative proportion deviated
0	$Z = 1$	4.0	41%
	$Z = 0$	-6.5	41%
1	$Z = 1$	5.0	21%
	$Z = 0$	-7.5	20%
2	$Z = 1$	6.0	9%
	$Z = 0$	-8.5	8%

all scenarios, is consistent with no confounding for the effect of Z on survival, as illustrated in Fig. 2 by the absence of any *open backdoor paths* (open paths consisting of arrows going in different directions and, therefore, non-causal paths) [9] connecting the treatment arm indicator Z and the future outcome Y_{t+1}. As a result, and because of the absence of loss to follow-up, a simple comparison of the estimated risks (i.e., cumulative incidences) in arm $Z = 1$ versus arm $Z = 0$ is an unbiased estimator of the intention-to-treat effect Pr$[Y_{t+1} = 1|Z = 1]$ versus Pr$[Y_{t+1} = 1|Z = 0]$ at any post-randomization time $t + 1 = 1, \ldots, 60$.

We are able to recover the true intention-to-treat effect in our study, regardless of the presence of protocol deviations, because unbiased estimation of the intention-to-treat effect only relies on the random assignment of Z and no loss to follow-up. In contrast, unbiased estimation of the *per-protocol effect* requires additional assumptions.

Defining and estimating the per-protocol effect

Let $Y_{t+1}^{\overline{a}=\overline{1}}$ denote an individual's indicator of death by month $t + 1$, had she, possibly contrary to fact, continuously followed the protocol in arm $Z = 1$. Similarly, let $Y_{t+1}^{\overline{a}=\overline{0}}$ denote this outcome by month $t+1$, had she, instead, continuously followed the protocol in arm $Z = 0$. We can then formally define the per-protocol effect at month $t+1$ as a contrast of the *counterfactual* risks:

$$\Pr\left[Y_{t+1}^{\overline{a}=\overline{1}} = 1|Z = 1\right] \text{ versus } \Pr\left[Y_{t+1}^{\overline{a}=\overline{0}} = 1|Z = 0\right]. \quad (3)$$

Note that, because Z was randomly assigned, we could alternatively define the per-protocol contrast as Pr$\left[Y_{t+1}^{\overline{a}=\overline{1}} = 1\right]$ versus Pr$\left[Y_{t+1}^{\overline{a}=\overline{0}} = 1\right]$ (unconditional on Z). Many randomized trials include a "naive" per-protocol analysis in which the survival curves are estimated after censoring participants at the time that they deviate from the protocol. This "naive" approach generally fails to recover the true per-protocol effect because it fails to account for confounding for the effect of received treat-

ment due to risk factors that affect both future adherence and survival. In Fig. 2, such confounding is represented by open backdoor paths connecting A_{t-1} and A_t to Y_{t+1}, e.g., the path $A_t \leftarrow L_t \leftarrow U \rightarrow Y_{t+1}$. The data-generating models we have described previously ensure the presence of this path by the dependence of A_t on past values of the time-varying risk factors $(L_0, \ldots, L_t)$, the dependence of L_t on U, and the dependence of Y_{t+1} on U. As described in the section "Data-generating models", we varied the degree of confounding (strong, moderate, or weak) by varying the magnitude of the parameter θ_1 in the model (2), which quantifies the strength of the dependence of Y_{t+1} on U.

Even though there is confounding for the per-protocol effect, the data generation mechanism in our study still allows unbiased estimation of the per-protocol effect as long as the study actually recorded all monthly covariates L_t and treatments A_t. Graphically, in Fig. 2 there are no open backdoor paths connecting A_{t-1} and A_t to Y_{t+1} *conditional* on past time-varying covariate changes [9]. For example, the open backdoor path $A_t \leftarrow L_t \leftarrow U \rightarrow Y_{t+1}$ is blocked by conditioning on L_t. Note that the measurement of the variable U is unnecessary to adjust for confounding when the variables L_t are measured in all t.

However, valid estimation of the per-protocol effect (3) requires the use of adjustment methods that, like IP weighting, can handle the fact that L_t is affected by past treatment [3, 4, 10]. We give a detailed description of the IP weighting algorithm in Additional file 2 and the R code in Additional file 1. Briefly, this approach involves: (1) as in the naive analysis, censoring participants when they deviate from their assigned protocol; (2) estimating IP weights which, at each time, are either 0 for censored participants or the reciprocal of the cumulative product of the time-varying probabilities of adherence to the protocol given the participant's measured confounder history up to that time for uncensored participants; and (3) estimating IP weighted survival curves. Risk differences and risk ratios can then be estimated by the complement of the IP weighted survival estimates. In addition to full measurement of the time-varying covariates, the validity of this approach also relies on correct specification of the model for the adherence probabilities in step 2.

Estimating the per-protocol effect under interval measurement

In practice, many randomized trials are conducted as interval cohorts such that adherence and covariates are recorded only at regular, scheduled follow-up times. When there are gaps between measurement times, the full history of treatment and covariate changes over the follow-up will not be completely observed and, generally, there will be unmeasured confounding; that is, under our data-generating assumption represented by Fig. 2, open backdoor paths will remain after conditioning on only the measured past. Also, the full history of treatment changes will be only partially observed. Under a non-null scenario, failure to measure interim treatment changes may produce an additional source of unmeasured confounding for treatment effects even at measured times; e.g., in Fig. 2, were there an arrow from A_{t-1} into Y_{t+1}, then an unblockable open backdoor path (by failure to measure A_{t-1}) connecting A_t and Y_{t+1} would be present. Partial knowledge of treatment changes thus also requires some form of imputation to estimate the per-protocol effect which is defined by counterfactual intervention in all months, not only months in which measurements are taken. Any imputation method may rest on strong assumptions, for example, imputation under the assumption that treatment does not change during measurement gaps or under missing at random (MAR) assumptions [11].

Suppose, without loss of generality, that the interval between measurements is constant throughout the follow-up, e.g., m months. We computed an IP weighted estimator of the per-protocol effect (3) and corresponding estimates of the counterfactual survival curves had all participants continuously complied with the protocol in each treatment arm under an interval-cohort scenario with $m = 12$, that is, a scenario in which treatment and covariate changes are measured only at baseline and then every 12 months. In interim months, treatment and covariates were set to the last measured value and the contribution to the weight cumulative product set to 1 for all subjects at these times. In this scenario, there will be residual confounding by failure to adjust for time-varying covariates at unmeasured times. At measured times, IP weights can only be based on the inverse probability that a subject continues to adhere in month s given her partially measured confounder history. This probability is unknown under our data-generating mechanism (because we generated each A_t from the full history). Thus, we would also expect some bias due to model misspecification under this scenario. Here we chose to model adherence based on the cumulative average of past measured values of the continuous time-varying covariate (based on only the baseline and every 12-month measurement) and the current value of the binary covariate (as the value from the previous month, the true value needed, will not be measured in this case).

Results

Intention-to-treat effect estimates

Figure 3 shows the estimated intention-to-treat survival curves $\Pr[Y_{t+1} = 0 | Z = 1]$ and $\Pr[Y_{t+1} = 0 | Z = 0]$ based on the the cumulative proportion of deaths in each arm by each follow-up month. Results are shown for the "strong confounding" scenario and the main study of approximately 40% deviators per arm (Scenario 0 in
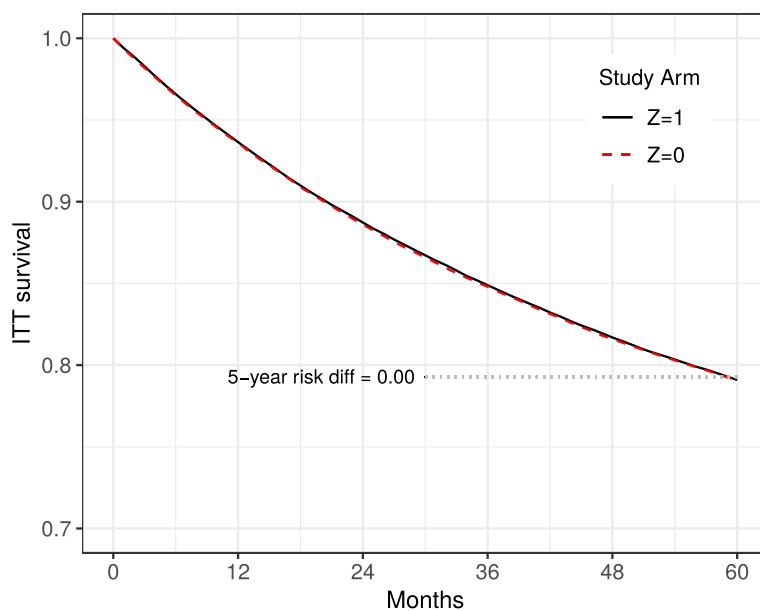
Fig. 3 Intention-to-treat survival estimates by treatment arm

Table 1). As expected, there is no bias in these estimates of the intention-to-treat effect; the curves completely overlap, which is consistent with the fact that the true intention-to-treat effect is null in all months $t + 1$.

Naive versus IP weighted per-protocol effect estimates under full measurement

As illustrated by the top panel of Fig. 4, in our study a "naive" unweighted estimator that ignores time-varying confounders fails to recover the true null per-protocol effect because the curves do not overlap. Rather the estimates of the per-protocol 5-year risk difference and risk ratio for standard versus new treatment are 0.11 and 1.77, respectively. The bottom panel of Fig. 4 shows IP weighted estimates of the per-protocol effect under full measurement of the time-varying covariates ($m = 0$). As expected, the estimated survival curves completely overlap, consistent with the truth, which is null. Figure 4 depicts results only under strong confounding. As expected, survival estimates across treatment arms under the naive approach that ignores confounding become closer as the strength of confounding weakens, while IP weighted estimates of the survival curves completely overlap under all scenarios (weak and moderate results are not shown).

IP weighted per-protocol effect estimates under interval measurement

In the interval-measurement scenario, we are generally unable to recover the truth of no per-protocol effect. In our study, IP weighted per-protocol effect estimates

under $m = 12$ diverged from the null as the strength of confounding increased. Specifically, Fig. 5 shows that differences in the survival curves increase with the strength of confounding, which results in 5-year risk difference/risk ratio estimates of 0.034/1.19 under strong confounding, 0.028/1.12 under moderate confounding, and 0.01/1.03 under weak confounding in our large sample.

Figure 6 illustrates that, even under strong confounding, bias decreases with more frequent measurement; estimates of the 5-year risk difference get closer to the truth of zero with decreasing m. Specifically, the IP weighted estimates of the risk difference/risk ratio were 0.017/1.02 under $m = 3$, 0.029/1.04 under $m = 6$, and 0.034/1.19 under $m = 12$.

Finally, Fig. 7 illustrates that, even under strong confounding and long interval measurement ($m = 12$), bias diminishes with decreasing non-adherence. Specifically, when the proportion of deviators decreased from approximately 40% (Scenario 0 in Table 1) to 20% (Scenario 1 in Table 1), the IP weighted estimates of the risk difference/risk ratio were closer to the null. Bias was negligible, with risk difference/ratio estimates of 0.004/1.005, when there were fewer than 10% deviators per arm (Scenario 2 in Table 1).

Discussion

We used a simulation to study bias in the estimation of per-protocol effects in randomized trials with interval-cohort designs. Bias arose even using methods such as
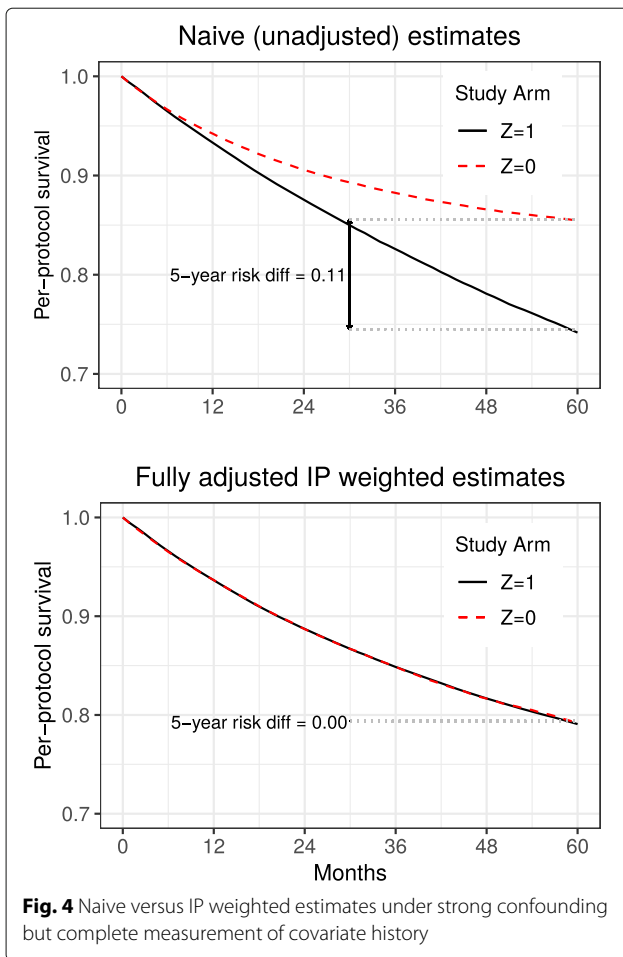
Fig. 4 Naive versus IP weighted estimates under strong confounding but complete measurement of covariate history
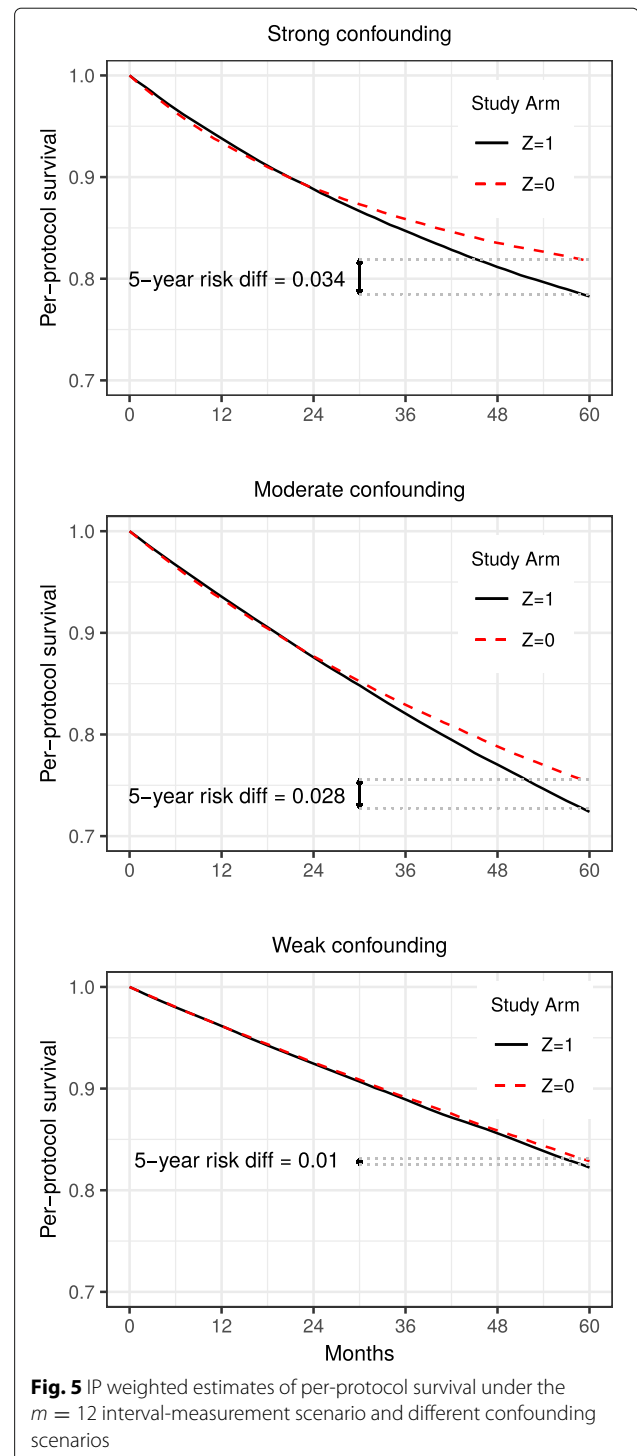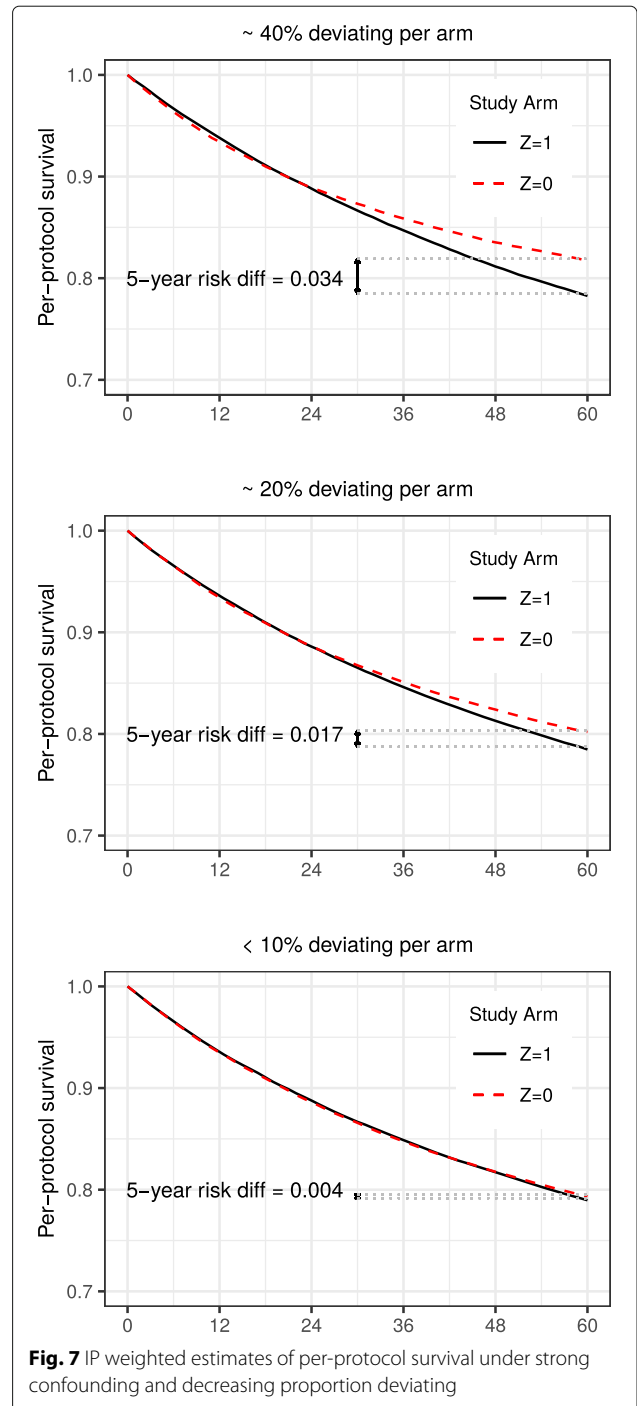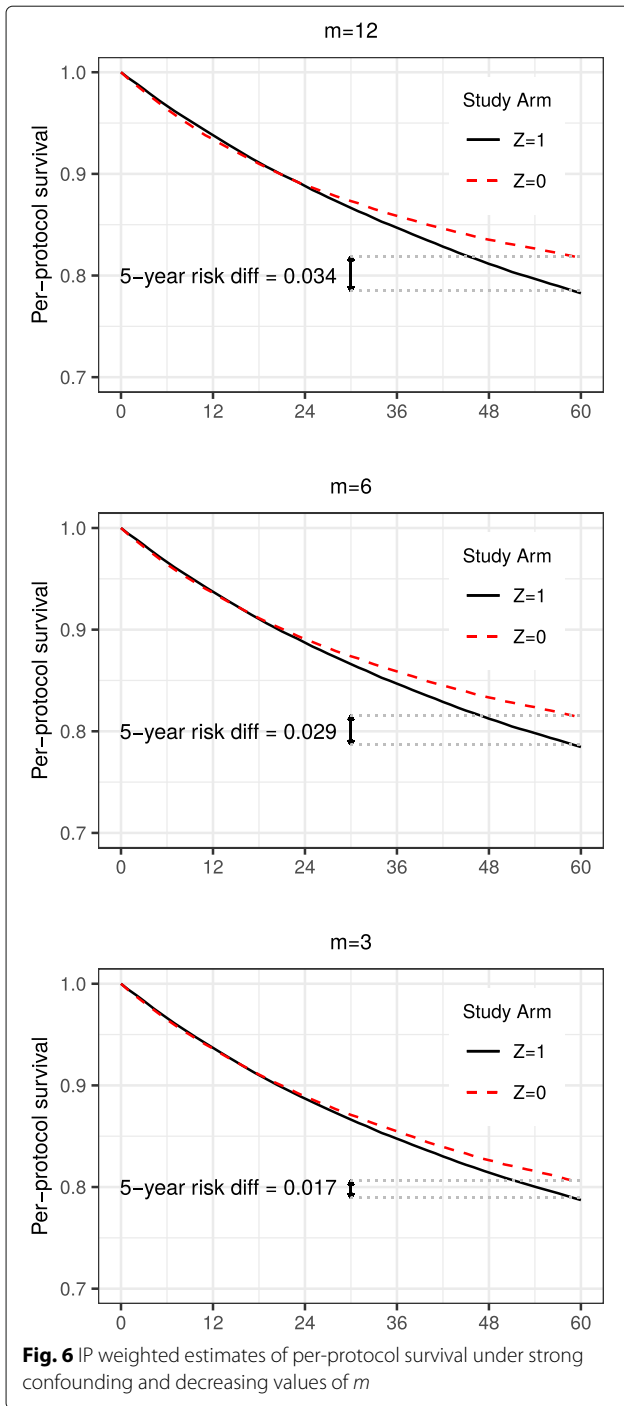

Fig. 5 IP weighted estimates of per-protocol survival under the $m = 12$ interval-measurement scenario and different confounding scenarios

IP weighting, which appropriately adjust for time-varying confounders. However, IP weighted estimates were less biased than estimates from a naive analysis that ignored time-varying confounding.

We considered the simple case of per-protocol effects defined by static treatment strategies (e.g., always take the new treatment versus always take the standard treatment), but our approach could also be applied to dynamic strategies under which treatment changes in response to pre-specified events (e.g., a drug toxicity) [12–14]. Also, we considered a simulation without censoring by loss to follow-up. Censoring may prevent unbiased estimation of both per-protocol and intention-to-treat effects without sufficient and appropriate adjustment for baseline and time-varying covariates [10, 15].

The bias created by interval measurement in the estimation of time-varying treatment effects has been previously highlighted in the computer science literature [16] and in epidemiological studies such as the Framingham Heart Study and the Nurses' Health Study [7, 17]. In practice, the interval length required to make the

bias negligible will depend on the frequency with which treatment and confounders can change. For example, in studies of treatments that rarely change more than once per month (like the one in our simulation), an interval length of one month will likely suffice. In other studies, measures of more frequent covariate changes may be

Fig. 6 IP weighted estimates of per-protocol survival under strong confounding and decreasing values of *m*

Fig. 7 IP weighted estimates of per-protocol survival under strong confounding and decreasing proportion deviating

necessary. In addition to more frequent in-person follow-up, complementary data sources such as electronic health records and pill cap monitors can help capture these changes.

Conclusions

The bias that arises from interval measurement highlights the need for randomized trials designed to collect post-baseline data on time-varying prognostic factors and adherence. This data may be obtained from various sources (e.g., more frequent in-person follow-up, electronic health records, pill cap monitors). Such planning, aided by the use of causal diagrams representing subject matter knowledge and assumptions, will ultimately provide improved estimates of the per-protocol effect, an informative complement to the intention-to-treat effect.

Additional files

Author details
[1]Department of Population Medicine, Harvard Medical School & Harvard Pilgrim Health Care Institute, Boston, USA. [2]Pathways M.D. Program, Harvard Medical School, Boston, USA. [3]Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, USA. [4]Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, USA. [5]Harvard-MIT Division of Health Sciences and Technology, Boston, USA. [6]Department of Epidemiology, Boston University School of Public Health, Boston, USA.

References

1. Hernán MA, Robins JM. Per-protocol analyses of pragmatic trials. N Engl J Med. 2017;14:1391–8.
2. Hernán MA, Hernández-Díaz S. Beyond the intention-to-treat in comparative effectiveness research. Clin Trials. 2012;1:48–55.
3. Robins JM. A new approach to causal inference in mortality studies with a sustained exposure period: application to the healthy worker survivor effect. Math Model. 1986;7:1393–512.
4. Robins JM. Addendum to "A new approach to causal inference in mortality studies with a sustained exposure period: application to the healthy worker survivor effect". Comput Math Appl. 1987;14:923–45.
5. Robins JM. Health service research methodology: a focus on AIDS. In: Sechrest L, Freeman H, Mulley A, editors. Washington, DC: US Public Health Service, National Center for Health Services Research; 1989. p. 113–59.
6. Robins JM. Correction for non-compliance in equivalence trials. Stat Med. 1998;17:269–302.
7. Hernán MA, McAdams M, McGrath N, Lanoy E, Costagliola D. Observation plans in longitudinal studies with time-varying treatments. Stat Methods Med Res. 2009;18(1):27–52.
8. Robins JM, Finkelstein D. Correcting for non-compliance and dependent censoring in an AIDS clinical trial with inverse probability of censoring weighted (IPCW) log-rank tests. Biometrics. 2000;56(3):779–88.
9. Pearl J. Causal diagrams for empirical research. Biometrika. 1995;82: 669–710.
10. Toh S, Hernán MA. Causal inference from longitudinal studies with baseline randomization. Int J Biostat. 2008;4(1):22.
11. Little RJA, Rubin DB. Statistical analysis with missing data. New York: John Wiley & Sons; 2002.
12. Hernán MA, Lanoy E, Costagliola D, Robins JM. Comparison of dynamic treatment regimes via inverse probability weighting. Basic & Clin Pharmacol & Toxicol. 2006;98:237–42.
13. Orellana L, Rotnitzky A, Robins JM. Dynamic regime marginal structural mean models for estimation of optimal dynamic treatment regimes, Part I: Main content. Int J Biostat. 2010;6:Article 7.
14. Orellana L, Rotnitzky A, Robins JM. Dynamic regime marginal structural mean models for estimation of optimal dynamic treatment regimes, Part II: Proofs and additional results. Int J Biostat. 2010;6:Article 8.
15. Little RJ, D'Agostino R, Cohen ML, Dickersin K, Emerson SS, Farrar JT, Frangakis C, Hogan JW, Molenberghs G, Murphy SA, Neaton JD, Rotnitzky A, Scharfstein D, Shih WJ, Siegel JP, Stern H. The prevention and treatment of missing data in clinical trials. N Eng J Med. 2012;367(14): 1355–60.
16. Schulam P, Saria S. Discretizing Logged Interaction data biases learning for decision-making; 2018. (pre-print) https://arxiv.org/abs/1810.03025.
17. Robins JM, Hernán MA, Siebert U. Effects of multiple interventions. In: Ezzati M, Lopez AD, Rodgers A, Murray CJL, editors. Comparative quantification of health risks: global and regional burden of disease attributable to selected major risk factors. Geneva: World Health Organization; 2004. p. 2191–230.

Publisher's Note