



Deposited via The University of Sheffield.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/id/eprint/236930/>

Version: Published Version

Article:

Campbell, H., Latimer, N., Jansen, J.P. et al. (2025) Augmented two-stage estimation for treatment switching in oncology trials: Leveraging external data for improved precision. *Statistical Methods in Medical Research*, 34 (12). pp. 2249-2269. ISSN: 0962-2802

<https://doi.org/10.1177/09622802251374838>

Reuse

This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here:

<https://creativecommons.org/licenses/>

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

Augmented two-stage estimation for treatment switching in oncology trials: Leveraging external data for improved precision

Statistical Methods in Medical Research

2025, Vol. 34(12) 2249–2269

© The Author(s) 2025



Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/09622802251374838

journals.sagepub.com/home/smm

Harlan Campbell^{1,2} , Nicholas Latimer³ ,
Jeroen P Jansen^{1,4} and Shannon Cope¹

Abstract

Randomized controlled trials in oncology often allow control group participants to switch to experimental treatments, a practice that, while often ethically necessary, complicates the accurate estimation of long-term treatment effects. When switching rates are high or sample sizes are limited, commonly used methods for treatment switching adjustment (such as the rank-preserving structural failure time model, inverse probability of censoring weights, and two-stage estimation) may produce imprecise estimates. Real-world data can be used to develop an external control arm for the randomized controlled trial, although this approach ignores evidence from trial subjects who did not switch and ignores evidence from the data obtained prior to switching for those subjects who did. This article introduces “augmented two-stage estimation” (ATSE), a method that combines data from non-switching participants in a randomized controlled trial with an external dataset, forming a “hybrid non-switching arm”. While aiming for more precise estimation, the augmented two-stage estimation requires strong assumptions. Namely, conditional on all the observed covariates: (1) a participant’s decision to switch treatments must be independent of their post-progression survival, and (2) individuals from the randomized controlled trial and the external cohort must be exchangeable. With a simulation study, we evaluate the augmented two-stage estimation method’s performance compared to two-stage estimation adjustment and an external control arm approach. Results indicate that performance is dependent on scenario characteristics, but when unconfounded external data are available, augmented two-stage estimation may result in less bias and improved precision compared to two-stage estimation and external control arm approaches. When external data are affected by unmeasured confounding, augmented two-stage estimation becomes prone to bias, but to a lesser extent compared to an external control arm approach.

Keywords

Evidence synthesis, survival analysis, comparative effectiveness, treatment switching, treatment crossover, health technology assessment, time-to-event outcomes

1 Introduction

Randomized controlled trials (RCTs) in clinical research often include the option for participants randomized to the control treatment (i.e. placebo or standard of care) to switch/crossover to the experimental treatment after a predefined time. This practice appears increasingly common in oncology and, from an ethical perspective, helps ensure that trial participants are

¹Precision AQ, HEOR - Evidence Synthesis and Decision Modeling, Vancouver, BC, Canada

²Department of Statistics, University of British Columbia, Vancouver, BC, Canada

³School of Health and Related Research, University of Sheffield, Sheffield, UK

⁴School of Pharmacy, University of California San Francisco, San Francisco, CA, USA

Corresponding author:

Harlan Campbell, Department of Statistics, University of British Columbia, Earth Sciences Building, 2207 Main Mall, Room 3182, Vancouver, BC V6T 1Z4, Canada.

Email: harlan.campbell@precisionaq.com

not denied access to potentially beneficial new treatments. Allowing for treatment switching (also known as “crossover”) is also thought to improve recruitment/enrolment,^{1,2} (but see Chen and Prasad³ who conclude otherwise). Unfortunately, treatment switching can complicate the estimation of overall survival (OS) benefits, as decision-makers are often interested in the hypothetical treatment effect of the experimental treatment versus the control treatment in jurisdictions where the experimental therapy is not yet available in the treatment pathway as a later line of therapy.⁴

In oncology trials, where progression-free survival (PFS) is frequently the primary endpoint, treatment switching in the control group upon disease progression can obscure the estimation of OS—a key measure for reimbursement and health technology assessments (HTAs). To address these challenges, statistical adjustment methods have been proposed, such as rank-preserving structural failure time (RPSFT) models,⁵ inverse probability of censoring (IPCW),⁶ and two-stage estimation (TSE).⁷ These approaches improve upon simple adjustment methods that simply censor subjects who switch treatments, given that switchers often have a different prognosis than non-switchers. However, these adjustment methods for treatment switching often result in highly uncertain estimates, especially in studies with a high rate of switching and/or a small sample size.

The National Institute for Health and Care Excellence (NICE) Decision Support Unit recently published a Technical Support Document (TSD 24, April 2024) that summarizes recommendations for adjusting survival time estimates in the presence of treatment switching in clinical trials. TSD 24 highlights the potential of using external data for treatment switching adjustment following similar calls in earlier work.^{8–10} However, until now, external data has only been used for treatment switching adjustment in select examples, where external evidence was used to (1) construct an external control arm (ECA) that was not impacted by switching and replace the RCT control arm considering alignment with the target trial population¹¹; (2) validate and select between different treatment switching adjustment methods (IPCW, TSE, and RPSFTM)¹²; (3) estimate what post-progression survival (PPS) would have been in the RCTs had treatment switching not occurred.¹³ This final example was the first to integrate external data into a TSE treatment switching adjustment model.¹³ Although the Evidence Review Group report¹⁴ raised concerns with regards to the use of external data in this case. Latimer and Abrams¹⁵ noted that “the deliberations of the [NICE] Appraisal Committee regarding TA171 demonstrated openness to the use of external data in the presence of treatment switching.”

In this article, we propose an alternative way of using external data for treatment switching adjustment based on the concept of a “hybrid control arm.” Methods for hybrid-control arm studies are not new, but are only recently being considered in oncology research^{16,17} (e.g. using a control arm formed from historical clinical trials in metastatic colorectal cancer¹⁸ or from EHR-derived data, such as the Flatiron Health database for advanced non-squamous non-small cell lung cancer (aNSCLC)¹⁹). Our proposed augmented two-stage estimation (ATSE) method leverages a “hybrid non-switching arm” for comparison with the RCT switching arm, enabling more efficient estimation of survival beyond the switching time-point. We detail the ATSE method in Section 2, present a simulation study in Section 3, and conclude with a discussion in Section 4.

2 Methods

We begin by defining basic notation and summarizing the standard two-step estimation (TSE) procedure. We then describe the augmented TSE (ATSE), which uses a hybrid non-switching arm to estimate the effect of treatment switching on survival. Note that adjustment for treatment switching is typically needed when the switching patterns observed in the trial do not represent standard clinical practice; see TSD 24 for details. Also, treatment switching can take many forms (e.g. switches from the experimental group onto the control treatment, and switches from either randomized group onto other treatments). We focus exclusively on switches from the control group to the experimental treatment that do not represent standard clinical practice.

2.1 Notation

Let us begin by defining some useful notation; see Table 1 for a summary overview. Suppose subjects in the RCT are randomly allocated to either the “control arm” or the “treatment arm,” with the primary endpoint of interest being right-censored OS. Subjects who are randomized to the control are allowed to switch from the control to the experimental treatment following observed/confirmed disease progression, and we assume that any switches are likely to happen immediately or shortly following progression. Suppose also that patient-level data from an external cohort is available for subjects treated with the control who did not switch onto the experimental treatment. Finally, suppose a decision-maker is interested in evaluating the hypothetical OS benefit of the treatment versus the control in a jurisdiction where the experimental treatment is not available in subsequent lines of therapy.

Table 1. Overview of notation.

Symbol	Description
S_i	Trial participation indicator for subject i ; $S_i = 1$ if subject i is in the RCT, $S_i = 0$ if in external data
$A_i = (A_{i1}, A_{i2})$	Binary treatment assignment sequence for subject i : A_{i1} for initial assignment, A_{i2} for post-switch
W_i	Treatment switch indicator for subject i ; $W_i = 1$ if switched to treatment, $W_i = 0$ otherwise
$T_i = (TTP_i, PPS_i, OS_i)$	Time-to-event outcomes for subject i : time to progression (TTP), post-progression survival (PPS), and overall survival (OS)
$C_i = (TTC_i, PPC_i, OSC_i)$	Indicators for TTP, PPS, and OS, respectively, for subject i , equal to 1 if observed (= 1), and equal to 0 if censored (= 0).
$AdmC_i$	Administrative censoring time
Z_i	Indicator if individual i was given (or would have hypothetically have been given) the option to switch ($Z_i = 1$, if yes; $Z_i = 0$, otherwise)
\mathbf{X}_i	Vector of covariates, for subject i , fixed at baseline or measured at progression
\mathbf{X}_{1i}	Subset of \mathbf{X}_i : common causes of post-progression survival (PPS_i), switching (W_i), and trial participation status (S_i)
\mathbf{X}_{2i}	Subset of \mathbf{X}_i : common causes of PPS and trial participation status (i.e. S_i)
α	Intercept parameter in the AFT model
μ	Parameter for the effect of switching in the AFT model
γ	Coefficient vector for covariates in the AFT model
ϵ_i	Random error term in AFT model (extreme value distribution)
σ	Scale parameter in Step 1 TSE AFT model
\hat{U}_i	Estimated counterfactual PPS time if the subject had not switched treatments
\widehat{OS}_i	Adjusted overall survival time = $PPS_i + \hat{U}_i$
β	Intercept in Step 1 AFT model for ATSE
ρ	Parameter measuring the difference between RCT and the external cohort in ATSE
η	Covariate effect vector in ATSE model (Step 1)
τ	Scale parameter in Step 1 ATSE AFT model
\hat{w}_i	Weight assigned to external data subject i , defined as a function of $\hat{\rho}$
c	Pre-specified decay factor controlling the influence of external data in ATSE

RCT: randomized control trial; TTP: time to progression; PPS: post-progression survival; OS: overall survival; AFT: accelerated failure time; TSE: two-stage estimation; ATSE: augmented two-stage estimation.

The i -th subject contributes S_i , $A_i = (A_{i1}, A_{i2})$, W_i , $T_i = (TTP_i, PPS_i, OS_i)$, $C_i = (TTC_i, PPC_i, OSC_i)$, $AdmC_i$, Z_i , and \mathbf{X}_i . Let S_i denote trial participation status, with $S_i = 1$ indicating that the i -th subject is in the RCT and $S_i = 0$ indicating that the i -th subject is in the external data. Let $A_i = (A_{i1}, A_{i2})$ correspond to the i -th subject's sequence of binary treatment assignments (with 0 corresponding to control and 1 corresponding to treatment). Let W_i be a binary indicator corresponding to a treatment switch (with $W_i = 1$ indicating a switch and $W_i = 0$ indicating no switch). For example, $A_i = (0, 1)$ implies $W_i = 1$ and indicates that the i -th subject was initially assigned/randomized to the control and then, following progression, switched to the experimental treatment. We assume that $S_i = 0$ implies that $A_i = (A_{i1}, A_{i2}) = (0, 0)$ since all subjects in the external data are "assigned" to the control and do not switch (switching is only an option for those in the RCT initially randomized to the control arm). Let $T_i = (TTP_i, PPS_i, OS_i)$ correspond to the i -th subject's time to progression (TTP), PPS, and OS times, with $C_i = (TTC_i, PPC_i, OSC_i)$ indicating if these are observed (= 1) or censored (= 0). An individual's post-progression survival will typically be calculated as the difference between their recorded OS and TTP (i.e. for the i -th subject, $PPS_i = OS_i - TTP_i$). Each individual also has an "administrative censoring time," $AdmC_i$, the timepoint at which they will be censored for administrative reasons regardless of their status in the study. This time typically corresponds to the overall study duration (e.g. study end date is 1 July 2000) relative to when a particular individual entered the study (e.g. individual i entered the study on 1 June 1998 and so $AdmC_i = 25$ months).

Let $Z_i = 1$ indicate that individual i was given (or would hypothetically have been given) the option to switch ($Z_i = 0$, otherwise). To be clear, if individual i was randomized to the control arm of the RCT (i.e. $S_i = 1$ and $A_{i1} = 0$), having a censored TTP (i.e. $TTC_i = 0$) or an entirely unobserved TTP (i.e. $TTP_i = 0$ and $TTC_i = 0$) would indicate that this individual had no option to switch because they were presumably being followed closely and progression was never observed, so $Z_i = 0$. However, if individual j was part of the external data (i.e. $S_j = 0$ and $A_{j1} = 0$), having a censored TTP (i.e. $TTC_i = 0$) or an entirely unobserved TTP (i.e. $TTP_i = 0$ and $TTC_i = 0$) might indicate that this patient was simply not being closely followed. It is important to think about the timepoint at which this patient would have hypothetically been given the option to switch to the experimental treatment had they been in the RCT and been followed more closely. If individual j 's time

of death is observed (i.e. $OSC_j = 1$) and we assume $Z_j = 1$, then, while their exact PPS is unknown, it will certainly be less than their time of death, and one could consider left-censoring (i.e. $PPS_j < (OS_j - TTP_j)$).

Finally, X_i corresponds to a vector of covariates measured at progression (or assumed to be fixed from baseline). Say X_i partitions as variables X_{1i} , which are common causes of PPS (PPS_i), switching (W_i), and trial participation status (S_i), and as variables X_{2i} , which are common causes of PPS (PPS_i) and trial participation status (S_i) (Table 1).

2.2 Two-stage estimation

As its name suggests, the TSE approach for treatment switching adjustment involves two main steps. First, one estimates the effect of treatment switching on PPS. Then, in the second step, this effect estimate is used to estimate counterfactual survival times that would have been observed if switching had not occurred.

Step 1. In the first step, a standard parametric accelerated failure time (AFT) model (such as a Weibull model) is fit to all subjects randomized to the control arm of the RCT (with observed progression times), relating PPS to switching (W), and all possible confounders (X_1) (i.e. all variables that are common causes for PPS and switching). For example, a Weibull model could be specified such that:

$$PPS_i = \exp(\alpha + \mu W_i + X_{1i}^T \gamma + \sigma \epsilon_i) \quad (1)$$

for the i -th subject (for all subjects for which $S_i = 1$ and $A_{i1} = 0$, and $Z_i = 1$), where α is the intercept parameter, μ is a parameter corresponding to the effect of switching, γ is a parameter-vector corresponding to the effect of the possible confounders, ϵ_i is an error term that has the extreme value distribution, and σ is the scale parameter. In an AFT model, the covariates act multiplicatively on time. For example, suppose the acceleration factor (AF) is $\exp(\mu) = 4$, then the effect of switching would be to quadruple the PPS.

Step 2. After obtaining parameter estimates from the AFT model, estimated counterfactual PPS times, \hat{U}_i are obtained as follows:

$$\hat{U}_i = W_i \frac{PPS_i}{\exp(\hat{\mu})} + (1 - W_i) PPS_i \quad (2)$$

for the i -th subject (for all subjects for which $S_i = 1$ and $A_{i1} = 0$, and $Z_i = 1$); where $\exp(\hat{\mu})$ is the estimated AF associated with switching obtained from the fitted AFT model in Step 1. Note that no adjustment is made for subjects who do not switch (i.e. if $W_i = 0$, then $\hat{U}_i = PPS_i$).

Suppose, for example, that the effect of switching is to double the PPS (i.e. $\exp(\hat{\mu}) = 2$). Then for subjects that switched, their counterfactual PPS times are obtained by halving their observed PPS times. Adjusted OS times are then obtained by adding the counterfactual PPS times to the observed TTP times: $\hat{OS}_i = TTP_i + \hat{U}_i$. If censoring is present, an additional step called “re-censoring” can be conducted. Since the TSE adjustment can result in informative censoring (if there is an association between switching and prognosis), the “re-censoring” step attempts to break the dependence between the counterfactual censoring time and switching. Briefly, the re-censoring step involves censoring all individuals in the RCT control group at the minimum of their administrative censoring time, $AdmC_i$, and $(AdmC_i / \exp(\hat{\mu}))$; see Zhang and Chen²⁰ for details, and see Latimer et al.²¹ for a discussion about the appropriateness of this step.

After estimating the untreated survival times for patients who switched treatments (and conducting any necessary re-censoring), a new “adjusted RCT” dataset is created. This dataset combines observed OS times for patients who did not switch treatments with adjusted OS times for those who did. This adjusted RCT dataset can then be used to estimate the relative treatment effect with respect to OS. For example, a Cox proportional hazards model could be fit to estimate a treatment switching-adjusted hazard ratio (HR). Alternatively, the relative treatment effect could be estimated non-parametrically as a treatment switching-adjusted difference in restricted mean survival time (dRMST). Valid confidence intervals for these estimates can be obtained by bootstrapping the entire adjustment and estimation process. To be clear, bootstrapping allows one to account for the additional uncertainty involved in estimating the AF in Step 1.

Note that to obtain unbiased results, the TSE method relies on the assumption of “no unmeasured confounders.” This means that the Step 1 AFT model must include all variables that predict both PPS and treatment switching. In other words, X_1 must include all prognostic variables that could influence a participant’s decision of whether to switch (or their doctor’s decision to recommend that they switch). It may be difficult to determine which variables influence a decision to switch, and ultimately, this assumption of “no unmeasured confounders” cannot be tested empirically. However, clinicians may provide

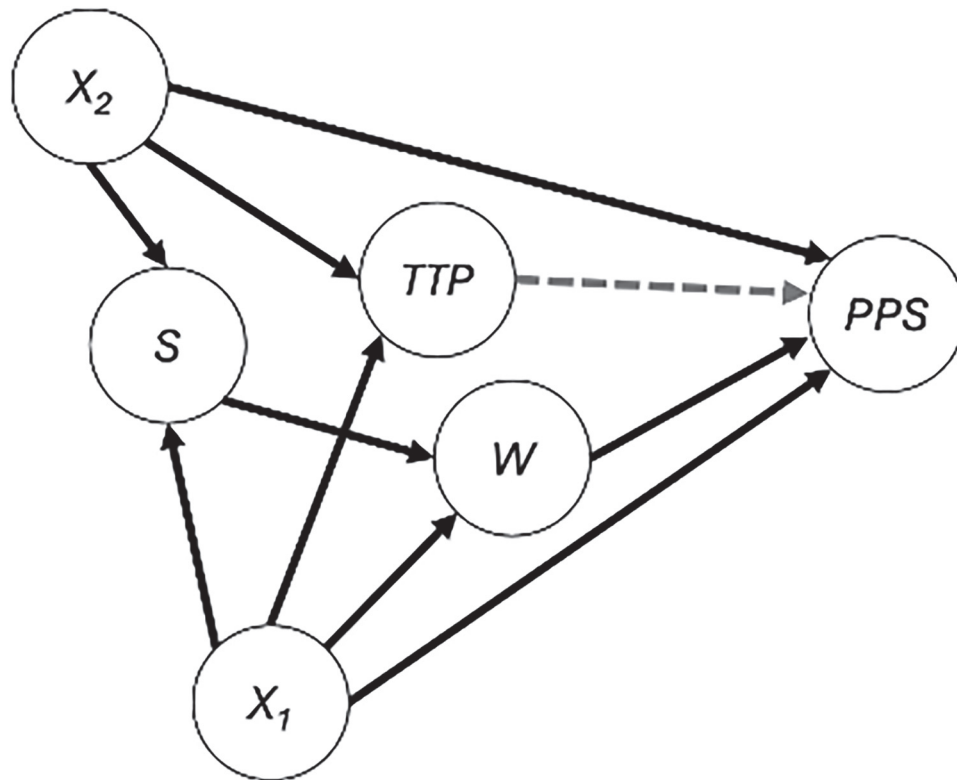


Figure 1. For estimating the effect of switching (W) on post-progression survival (PPS), there are two minimal sufficient adjustment sets: S, X_1 and X_1, X_2 . The AFT model fit in Step 1 of the TSE is restricted to only those individuals with $S = 1$ (i.e. only subjects in the RCT) and must therefore include X_1 (i.e. all variables that predict both PPS and treatment switching). The AFT model in Step 2 of the ATSE is fit to individuals from both the RCT ($S = 1$) and external data ($S = 0$) and must therefore include X_1 and X_2 (i.e. all variables that predict both PPS and treatment switching as well as all variables that predict both PPS and trial participation status). AFT: accelerated failure time; TSE: two-stage estimation; RCT: randomized controlled trial; ATSE: augmented two-stage estimation.

valuable information about treatment switching decisions and directed acyclic graphs (DAGs), such as the Figure 1 diagram, may be a useful tool to help inform covariate selection.²²

The DAG in Figure 1 can be understood to illustrate the causal relationships between each of the variables. For instance, the arrow from X_1 to W implies that the variables included in X_1 directly influence a decision to switch, and the arrow from X_1 to PPS implies that these variables also directly impact the PPS . For example, “disease severity” as measured shortly before progression might be included in X_1 and would influence the decision to switch (e.g. those with high severity are recommended by their doctor’s to not switch to the experimental treatment) and the duration of PPS .

2.3 Augmented TSE

For TSE, a limited sample size in the non-switching subjects in the control arm of an RCT increases uncertainty in the estimation of the AF (i.e. $\exp(\mu)$), leading to an imprecise estimate of the counterfactual survival times and ultimately the treatment switching-adjusted relative treatment effect. Leveraging external data for subjects assigned to the control treatment who do not switch (e.g. from historical RCTs or RWE) could reduce the uncertainty in the estimation of the AF, resulting in a more precise adjustment for treatment switching. If external data is available, the simplest approach might be to conduct external control arm study in which individuals from the external data source are selected (and typically weighted) to use as a stand-in control arm replacing the individuals randomized to control in the RCT; see Jaksa et al.²³ However, entirely ignoring the RCT control arm may be inefficient and selection of an appropriate, fit-for-use external cohort is critical to minimizing bias.

Various methods have been developed to construct the so-called “hybrid control arms,” whereby a trial’s small control arm is combined with individuals from an external data source. These methods are mostly Bayesian approaches and include (modified) power prior models (Ibrahim and Chen²⁴), commensurate prior models (Hobbs et al.²⁵), and robust meta-analytic predictive prior (RMAPP) models (Schmidli et al.²⁶). Most recently, the Bayesian latent exchangeability

prior (LEAP) model (Alt et al.²⁷) appears particularly promising (Campbell and Gustafson²⁸). However, since the TSE method is frequentist, for the ASTE we adopt the two-step dynamic borrowing approach recently proposed by Tan et al.,²⁹ which reflects a frequentist analog to the modified power prior method.

The augmented TSE method consists of the following four steps.

Step 1. In the first step, a standard parametric AFT model (such as a Weibull model) is fit comparing the PPS between those randomized to the control arm who did not switch and those in the external control arm, adjusting for covariates $X = X_1, X_2$. To be clear, one must adjust for all common causes of PPS and switching (i.e. adjust for X_1), as well as all common causes of PPS and trial participation status (i.e. adjust for X_2); see Figure 1.

For example, a Weibull model could be specified such that:

$$PPS_i = \exp(\beta + \rho S_i + X_i^T \eta + \tau \epsilon_i) \quad (3)$$

for the i -th subject (for all subjects for which $W_i = 0$, $A_{i1} = 0$, and $Z_i = 1$), where β is the intercept parameter, ϵ_i is an error term that has the extreme value distribution, and τ is the scale parameter. The value of the ρ parameter corresponds to the degree of dissimilarity between the two cohorts. Specifically, $\exp(\rho)$ is the AF associated with participation in the RCT versus the external cohort (after adjusting for all measured confounders). Suppose, for example, that $\exp(\rho) = 2$. This would indicate that individuals in the RCT have PPS times that are twice as long on average as individuals with similar values of X who are in the external data cohort. A large value of $|\rho|$ would therefore indicate a large degree of dissimilarity between cohorts that cannot be attributed to measured confounders and would suggest the possibility of substantial unmeasured confounding.

Dynamic borrowing then considers the degree of dissimilarity in determining how much information to borrow from the external cohort. If the magnitude of ρ is small, this suggests that the external data is relatively compatible with the RCT data. On the other hand, if the magnitude of ρ is large, this suggests that the external data is less compatible with the RCT data and should therefore be down-weighted in the analysis. To calculate the amount of cohort-level down-weighting as a function of ρ , Tan et al.²⁹ consider an exponential function which assigns each individual in the external data a weight equal to

$$\hat{w}_i = \exp(-c|\hat{\rho}|) \quad (4)$$

for the i -th subject (for all subjects for which $S_i = 0$); where $c > 0$ is a pre-specified constant “decay factor,” akin to the power parameter in the Bayesian modified power prior approach. A larger value of c will result in borrowing less information from the external data (i.e. in a faster decay to 0 as the difference between the RCT and external cohorts increases). All individuals in the RCT (all subjects for which $S_i = 1$) are given a weight equal to 1. In order to choose an appropriate value for c , the following example may be helpful. If the estimated AF, $\exp(\hat{\rho})$, is 1.2 (corresponding to a 20% increase in post-progression survival) and if $c = 4$, then each external data point will receive a weight of 0.48 when following equation (4) for the weights. In other words, the contribution of the external data will be approximately halved on account of the observed incompatibility.

To be clear, there are other options for the weight function defined in equation (4). Any function that is bounded between 0 and 1 and monotonically decreases weights with increasing $|\hat{\rho}|$ could be suitable. The key idea is to discount the external data's influence when the compatibility of the external data and the RCT data seems improbable according to the Step 1 AFT model.

Step 2. The second step of the ATSE method is to fit a second AFT model to the weighted control subjects (i.e. all subjects with $A_{i1} = 0$) with the weights defined from the previous step. This second AFT model relates PPS to switching (W), and includes covariates $X = X_1, X_2$, the same covariates which were included in the first step AFT model:

$$PPS_i = \exp(\alpha + \mu W_i + X_i^T \gamma + \sigma \epsilon_i) \quad (5)$$

for the i -th subject (for all subjects for which $A_{i1} = 0$, and $Z_i = 1$), where α is the intercept parameter, ϵ_i is an error term that has the extreme value distribution, and σ is the scale parameter. This is similar to the first step of the TSE approach, except that the dataset here is expanded to include the down-weighted external data, rather than only the control patients of the RCT. Note that adjusting for S_i in this model would effectively exclude the external data, and so S_i is absent.

Step 3. The third step of the ATSE approach is deriving the estimated counterfactual PPS times, \hat{U}_i , and OS times, \widehat{OS}_i (for all subjects with $S_i = 1$, $A_{i1} = 0$, and $Z_i = 1$). This is done as in TSE (see equation (2)) based on $\exp(\hat{\mu})$, the estimated AF associated with switching obtained from fitting model (5) in the second step. If censoring is present, “re-censoring” can be applied; see Latimer et al.²¹

Step 4. The fourth and final step of the ATSE approach is to estimate the relative treatment effect of interest based on a new “adjusted RCT” dataset, which combines observed OS times for RCT patients who did not switch treatments with adjusted OS times for those RCT patients who did switch. For instance, a Cox proportional hazards model, or an AFT model, relating OS to treatment at randomization (A_1) could be fit with the adjusted RCT dataset to estimate the relative treatment effect in terms of an HR or an AF. As with TSE, valid confidence intervals can be obtained by bootstrapping the entire adjustment and estimation process (which allows one to account for the uncertainty in estimating the weights in Step 1 and the AF in Step 2).

It is important to reiterate that, beyond the “no unmeasured confounders” assumption required in TSE (i.e. one must adjust for all variables that predict both PPS and treatment switching), the ATSE method requires an additional “no unmeasured confounders” assumption with respect to the external data. The DAG in Figure 1 illustrates that there are two minimal sufficient adjustment sets for estimating the effect of switching (W) on PPS: S, X_1 and X_1, X_2 . The AFT model fit in Step 1 of the TSE is restricted to only those individuals with $S = 1$ (i.e. only subjects in the RCT) and must therefore include X_1 (i.e., all variables that are common causes of PPS and W). The AFT model in Step 2 of the ATSE is fit to individuals from both the RCT ($S = 1$) and external data ($S = 0$) and must therefore include X_1 and X_2 . Down-weighting the external data according to the suspected amount of unmeasured confounding as determined by the AFT model in Step 1 of the ATSE will help reduce the impact of confounding bias, but cannot entirely eliminate it. With regards to covariate selection, a conservative approach would be to adjust for all covariates that are considered important prognostic factors.

We have included **TTP** in Figure 1 DAG with a dashed line towards **PPS** to indicate a potential direct causal link. Regardless of whether there is a direct causal link, **TTP** might serve as a useful “proxy confounder” in the sense that it might be highly correlated with X_1 and/or X_2 . For instance, if “disease severity” measured shortly before progression is deemed a confounder within X_1 or X_2 , but is unmeasured in the data, one might consider **TTP** as a proxy for this unmeasured confounder (since those with high “disease severity” will likely have shorter progression times and vice versa). VanderWeele³⁰ explains that “in most cases, adjustment for such a variable will reduce the bias due to confounding.” In order to adjust for TTP in the ATSE, one would simply expand the covariate set X to include **TTP** in Steps 1 and 2: $X = X_1, X_2, \text{TTP}$. In Section 3, we will consider the merits of this strategy.

3 Simulation study

3.1 Objectives

To evaluate the performance of the ATSE method, we conducted a simulation study designed to mirror typical settings in clinical research with treatment switching and incorporation of external data. This simulation study aimed to assess the accuracy and robustness of the ATSE approach under various conditions, including variations in switching rates, and different decay factors for the ATSE hybrid arm. Here, we provide a summary of the data-generating mechanism, the estimand of interest, and the methods we compared according to various performance measures. The simulation study was conducted using R. The code used to simulate the data is provided in the Supplementary Material.

3.2 Data generation and mechanism

We followed the data generation procedure used outlined by Latimer et al.³¹ for their eight “simple scenarios” with a few modifications. Specifically, we simulated RCT datasets with a sample size of $N_{RCT} = 500$ and 2:1 randomization in favor of the experimental group, and with treatment switching permitted from the control group to the experimental treatment following progression.

OS times were simulated based on a two-component mixture Weibull baseline survival function and were dependent on three binary variables: treatment, prognosis (*badprog*), and an unmeasured prognostic factor (U) (note the simulation study by Latimer et al.³¹ does not include a dependency on U , but is otherwise the same). The corresponding hazard function is

$$h_i(t) = h_0(t)\exp(\delta_1 trt_i + 0.3badprog_i - 0.3U_i) \quad (6)$$

for the i -th subject; where $h_0(t)$ represents the baseline hazard function and δ_1 represents the log HR (log-HR) associated with treatment. The bad prognosis and unmeasured confounder variables were simulated as independent binary variables

Table 2. Overview of simulated scenarios.

Scenario	Treatment effect	Switch proportion	Censoring	Sample size of the external control arm	True RMST for the control group
1	Low	Moderate	None	200	472.75
2	High	Moderate	None	200	472.75
3	Low	High	None	200	472.75
4	High	High	None	200	472.75
5	Low	Moderate	Moderate	200	368.60
6	High	Moderate	Moderate	200	368.60
7	Low	High	Moderate	200	368.60
8	High	High	Moderate	200	368.60
9	Low	Moderate	None	1000	472.75
10	High	Moderate	None	1000	472.75
11	Low	High	None	1000	472.75
12	High	High	None	1000	472.75
13	Low	Moderate	Moderate	1000	368.60
14	High	Moderate	Moderate	1000	368.60
15	Low	High	Moderate	1000	368.60
16	High	High	Moderate	1000	368.60

RMST: restricted mean survival time.

such that each individual had a 50% probability of a bad prognosis and a 50% probability of $U_i = 1$. We assumed no treatment effect heterogeneity (i.e. no effect modifiers/interaction terms). Also, while in reality certain covariates, such as prognosis, might change (worsen) over the course of the study (e.g. a subject's disease severity may be notably different at baseline than at progression), we assumed that all covariates remain fixed over the course of the study.

TTP times were simulated as a function of OS times, so that on average, an individual's TTP was one-third of their OS. More specifically, TTP times were equal to OS times multiplied by a random draw from a beta(5, 10) distribution. A short delay was assumed between an individual's true progression time and their observed progression time by setting the observed progression time as equal to their first "visit time" following the progression event, with "visits" simulated every 21 days from randomization to death.

We simulated external control datasets in the same way that the RCT datasets were simulated, but with $trt_i = 0$ for all subjects. Also, whereas individuals in the RCT had a 50% probability of a bad prognosis, individuals in the external control dataset had a 75% probability of a bad prognosis.

Sixteen scenarios were simulated varying (1) the magnitude of the treatment effect, (2) the degree of switching, (3) the censoring, and (4) the sample size of the external control arm; see Table 2. These scenarios were based on Latimer et al.'s³¹ eight simple scenarios. Specifically, for scenarios with "low treatment effect," we set $\delta_1 = -0.2$, whereas for scenarios "high treatment effect," we set $\delta_1 = -0.5$. For scenarios with "moderate switching," RCT individuals with poor prognosis had an 80% probability of switching, whereas those with a good prognosis group had a 30% probability of switching. For scenarios with "high switching," the probability of switching for those in the control arm was set to 90% in the poor prognosis group and to 60% in the good prognosis group. To be clear, in this simulation study, all individuals in the external control and all those in the experimental treatment arm of the RCT do not switch treatments. In the scenarios with "no censoring," individuals were only censored if still alive at 5000 days following randomization (the end of the study), whereas scenarios with "moderate censoring" censored all subjects at 546 days following randomization (the end of the study). Finally, the sample size of the external control arm was either $N_{EC} = 200$ (resembling historical trial data) or $N_{EC} = 1000$ (resembling registry data).

To investigate the impact of unmeasured confounding, for each of the sixteen scenarios, 2000 datasets were simulated with three different conditions. In Condition A ("Complete") there was no unmeasured confounding, meaning the unmeasured prognostic factor U did not impact the probability of switching and had the same distribution in the RCT and external control (i.e. all subjects in the RCT and external control, $\Pr(U_i = 1) = 0.5$). In Condition B ("Incomplete X_2 ") confounding bias in the external control was considered by setting the prognostic factor U in the RCT to $\Pr(U_i = 1) = 0.50$ and in the external control to $\Pr(U_i = 1) = 0.75$. Finally, in Condition C ("Incomplete X_1 "), confounding bias in the treatment switching-adjusted control arm of the RCT was considered, where the unmeasured prognostic factor U impacted switching, reducing the probability of switching by 20% for all individuals in the RCT for whom $U_i = 1$. In this case, the external data remains unbiased with $\Pr(U_i = 1) = 0.50$ for all subjects in both the RCT and the external data. Note that the same random

number seed was used for all three conditions within each scenario so that the RCT data simulated in Condition A (“Complete”) and Condition B (“Incomplete X_2 ”) will be identical and the external data simulated in Condition A (“Complete”) and Condition C (“Incomplete X_1 ”) will be identical.

3.3 Estimands of interest

As in Latimer et al.,³¹ the estimand of interest for the simulation study was the restricted mean survival time (RMST) in the control group at study end-date (5000 days for Scenarios 1–4 and 9–12; 546 days for Scenarios 5–8 and 13–16). The true RMST value in the control group was 472.75 days in Scenarios 1–4 and 9–12, and was 368.6 days in Scenarios 5–8 and 13–16, based on numerical integration.

When extrapolation was required to calculate the RMST, we used a flexible Royston-Parmar parametric splines model with three knots (using the RMST R package; <https://github.com/scientific-computing-solutions/RMST>), consistent with HTA recommendations.³²

3.4 Methods to be compared

We compared the following methods:

- *Oracle*: A standard ITT analysis undertaken on the simulated data prior to simulating the impact of switching, which represents the “truth” for each simulation.
- *ITT*: A standard ITT analysis on the simulated data after switching impacted the outcomes.
- *TSE*: Treatment switching-adjusted analysis using TSE based on only the RCT data with re-censoring applied (Section 2.2).
- *ECA*: A propensity-score-based analysis in which the external control data is used as an external control arm (and individuals randomized to control in the RCT are ignored). Specifically, we used the “weightit” function from the WeightIt R library³³ to derive the average treatment effect for the treated (ATT) weights using a logistic regression model. The weighted OS ECA data was then used to estimate the RMST in the control group.
- *ATSE* ($c = 1$): An analysis based on the proposed approach with decay factor set to $c = 1$ and re-censoring applied (Section 2.3). This smaller value of c will result in borrowing more information from the external data.
- *ATSE* ($c = 4$): An analysis based on the proposed approach with decay factor set to $c = 4$ and re-censoring applied (Section 2.3). This method considers a “mid-range” value of c .
- *ATSE* ($c = 8$): An analysis based on the proposed approach with decay factor set to $c = 8$ and re-censoring applied (Section 2.3). This larger value of c will result in borrowing less information from the external data.
- *ATSE* ($c = 4$) with adjustment for TTP: An analysis based on the proposed approach with decay factor set to $c = 4$ and re-censoring applied (Section 2.3), which includes TTP as a covariate in the ATSE Step 1 and Step 2 AFT models.

3.5 Performance measures

Consistent with Latimer et al.,³¹ the performance of each method was evaluated according to the percentage bias in the estimate of the control group RMST at the study end date. Percentage bias was estimated by taking the difference between the estimated RMST and the true RMST, expressed as a percentage of the true RMST. Root mean squared error (RMSE) and empirical standard errors (SEs) of the RMST estimates were also calculated for each method and expressed as percentages of the true RMST.

3.6 Results

Figure 2 plots the results for Scenario 1 and Figure 3 plots the results for Scenario 9. Figures 4 to 9 display results in nested loop plots for the percentage bias and empirical SE across all 16 scenarios. Tables 3 to 6 in the Appendix list the complete results. The simulation study provided very accurate estimates. For the mean RMST percentage bias, the Monte Carlo SE (MCSE) never exceeding 0.20% across all 16 scenarios. For the empirical SE as a percentage of true RMST, the maximum (across all 16 scenarios) MCSE was 0.14% and for the RMSE as a percentage of true RMST, the maximum MCSE was 0.34%. Error bars plotted in Figures 2 and 3 correspond to ± 1.96 times the MCSE. Finally, the differences between Figures 2 and 3 for the Oracle, ITT, and TSE methods provide an additional measure of the size of the Monte Carlo error (since, in the absence of Monte Carlo error, these should be identical).

Let us focus first on the results from Scenario 1. As expected, under all three conditions, the *ITT* analysis over-estimates the control group RMST, resulting in a percentage bias of between 4.4% and 5.4%. Under Condition A (“Complete”), all other methods (*TSE*, *ECA*, and *ATSE*) predict the control group RMST with negligible bias ($\leq 0.6\%$). Among these

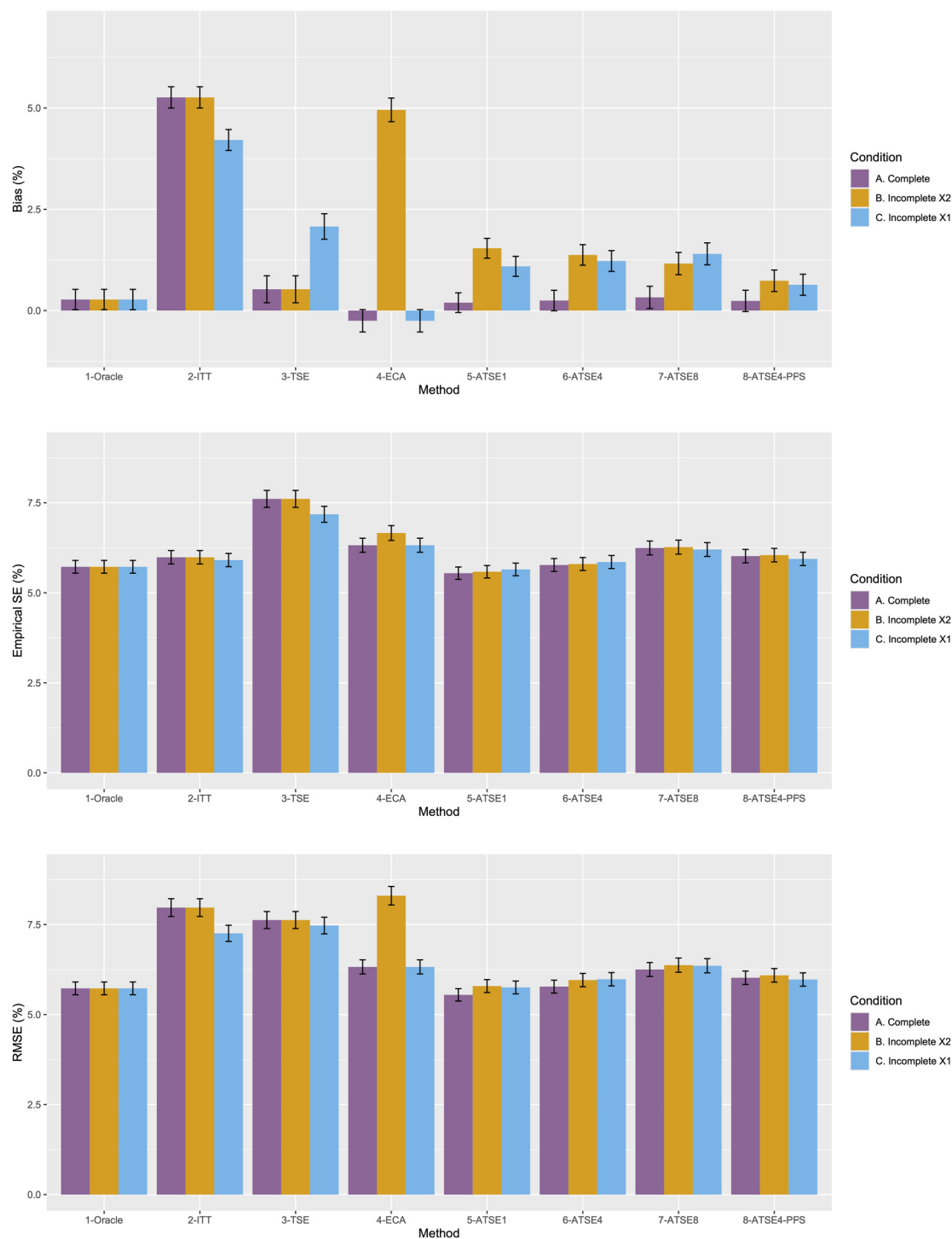


Figure 2. Simulation study results from Scenario I.

other methods, the *ASTE* analysis appeared to be the most efficient with SEs ranging from 5.4% to 6.2%. Notably, the *ECA* approach had an empirical standard error of 6.3%, lower than the *TSE*, which had an empirical standard error of 7.6%. It is interesting that the SE for the *ATSE* can be slightly lower than for the Oracle (compare 5.4% to 5.7%). This can be explained by the fact that more data is being used in the *ATSE* analysis. Under Condition B (“Incomplete X_2 ”), when there may be confounding bias with respect to the external control data, the *ECA* appears to be biased, over-estimating the control group RMST by 5.0%. The *ATSE* analyses are also biased, but to a much lesser degree, over-estimating the control group

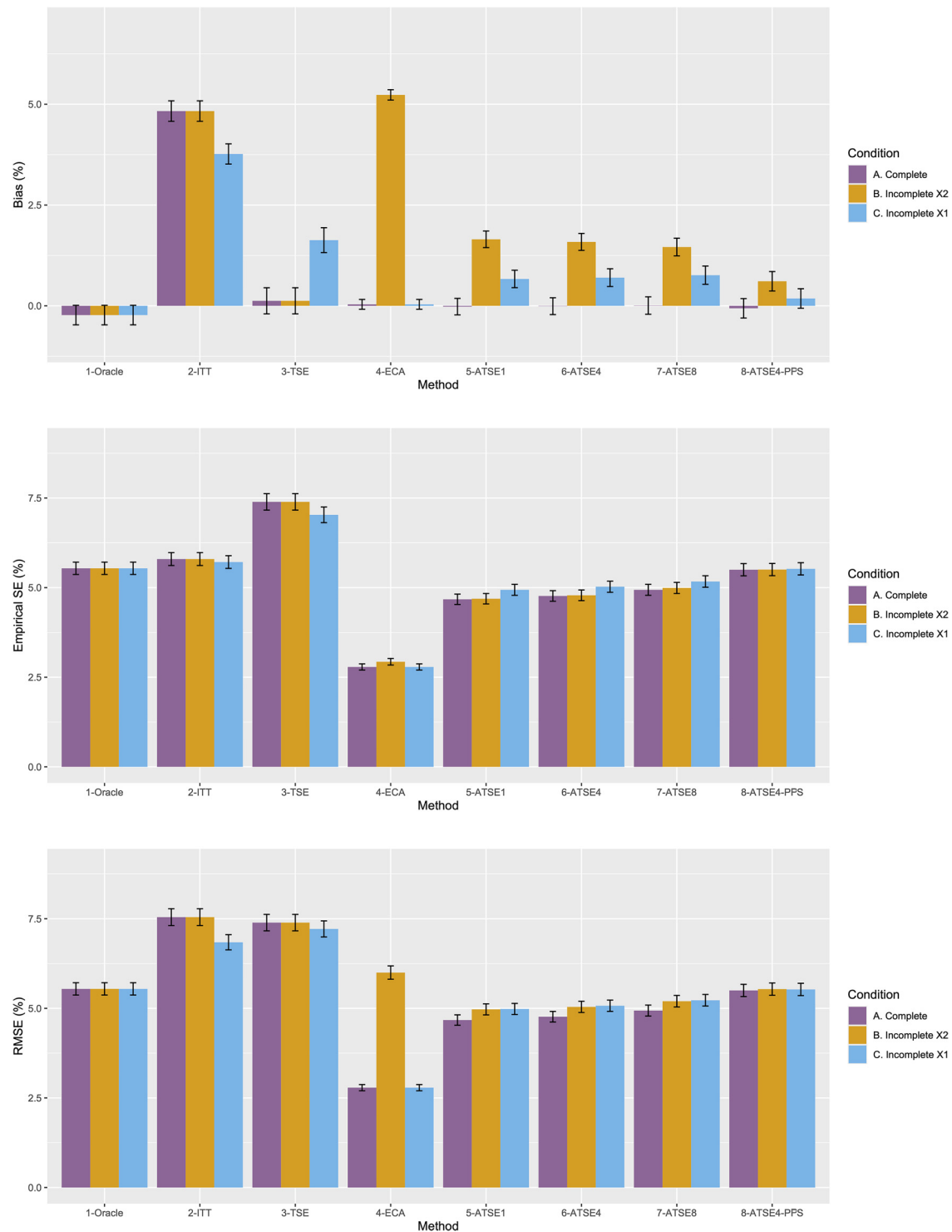


Figure 3. Simulation study results from Scenario 9.

RMST by only 1.2%–1.6%. Under Condition C (“Incomplete X_1 ”), when there may be confounding bias with respect to the treatment switching adjustment, the *TSE* appears to be biased, over-estimating the control group RMST by about 2.2%. The *ATSE* analyses are also biased, over-estimating the control group RMST by about 1.2%–1.5%. Results were similar across all scenarios.

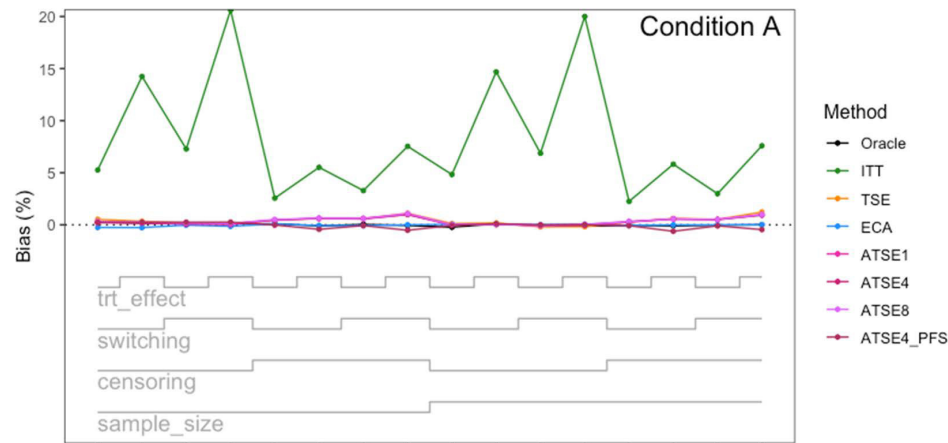


Figure 4. Simulation study results: Percentage bias for Condition A ("Complete") across the 16 scenarios.

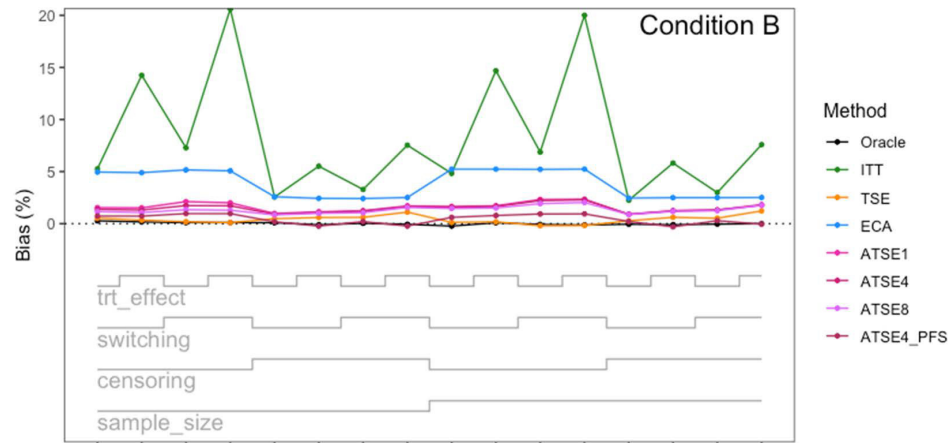


Figure 5. Simulation study results: Percentage bias for Condition B ("Incomplete X_2 ") across the 16 scenarios.

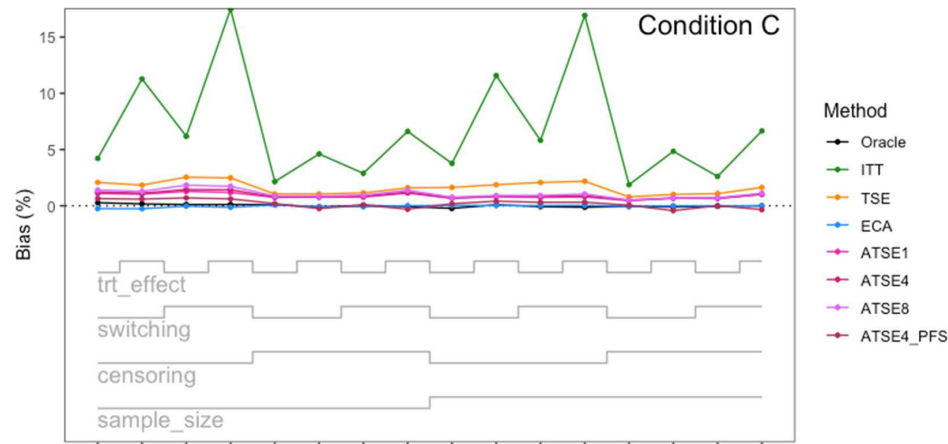


Figure 6. Simulation study results: Percentage bias for Condition C ("Incomplete X_1 ") across the 16 scenarios.

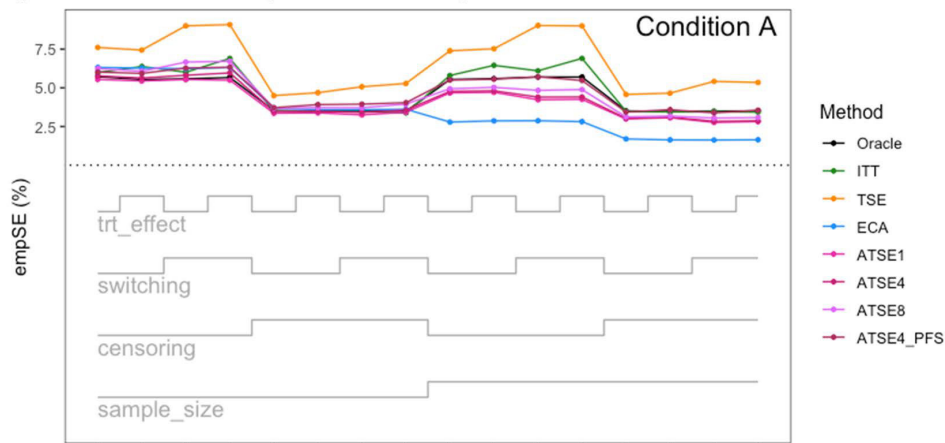


Figure 7. Simulation study results: Empirical SE for Condition A ("Complete") across the 16 scenarios.

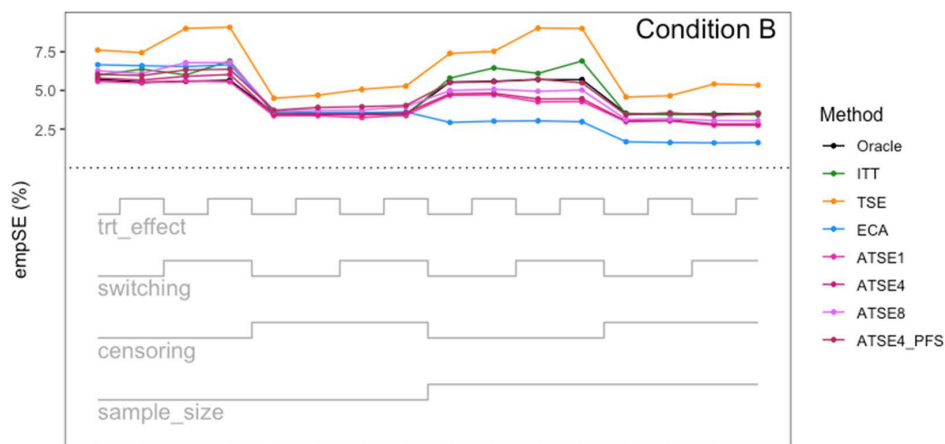


Figure 8. Simulation study results: Empirical SE for Condition B ("Incomplete X_2 ") across the 16 scenarios.

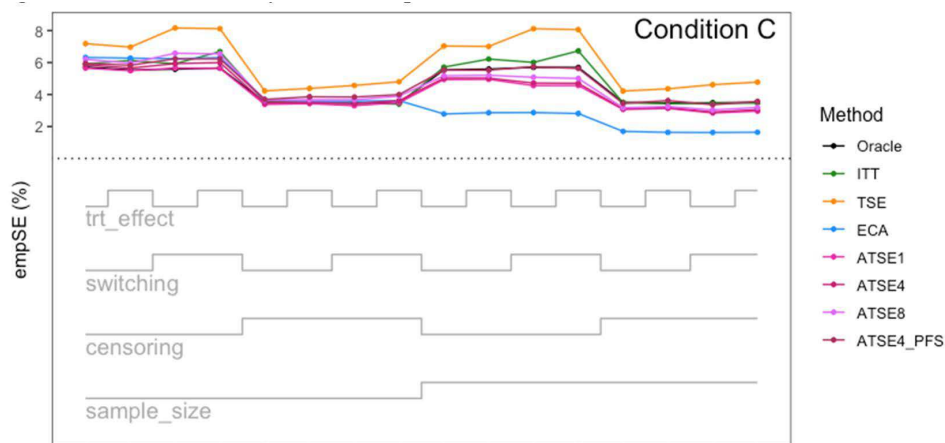


Figure 9. Simulation study results: Empirical SE for Condition C ("Incomplete X_1 ") across the 16 scenarios.

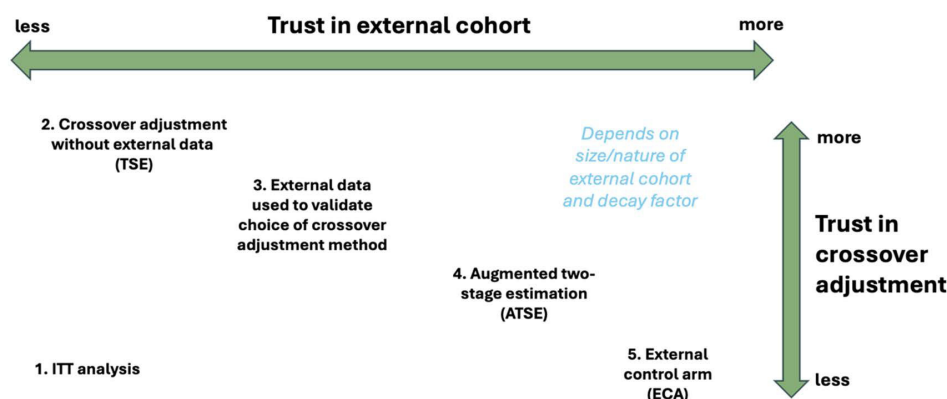


Figure 10. If the assumptions underlying the treatment switching adjustment are suspect, then the ECA may be preferable to the ATSE approach. Alternatively, if the assumption of exchangeability for the subjects in the external data is suspect, then the TSE may be preferable to the ATSE approach. If no assumptions can be relied upon, the ITT analysis (while biased) will be most appropriate. ECA: external control arm; ATSE: augmented two-stage estimation; TSE: two-stage estimation; ITT: intention-to-treat.

The impact of using different values for the decay factor with *ATSE* was, overall, rather minimal. Recall that a smaller decay factor corresponds to a higher degree of borrowing from the external data. We see that under Condition A (“Complete”), the empirical SE as a percentage of the true RMST is smallest when $c = 1$ and is highest when $c = 8$, meaning that more borrowing leads to more efficiency. Under Condition B (“Incomplete X_2 ”), where the external control data introduces bias, the percent bias in RMST is highest when $c = 1$, and is lowest when $c = 8$. In contrast, under Condition C (“Incomplete X_1 ”), the percent bias in RMST is lowest when $c = 1$, and is highest when $c = 8$. As such, the decay factor seems to correspond to a trade-off between the potential bias due to unmeasured confounding in either the RCT treatment switching-adjusted controls or the external control.

The impact of adjusting for TTP in the *ASTE* was notable. We see that the degree of bias due to confounding is substantially reduced at the cost of a modest increase in SE. We do not observe any substantial impact with respect to differing treatment effect sizes ($\delta_1 = -0.2$ vs. $\delta_1 = -0.5$), nor do we observe any substantial impact with respect to differing switching proportions or censoring. The *ASTE* method is consistently the most efficient approach when the sample size of the external control arm was $N_{EC} = 200$ (Scenarios 1–8). However, the ECA was notably superior when the sample size of the external control arm was $N_{EC} = 1000$ (Scenarios 9–16).

4 Conclusion

The TSD 24 specifically brings attention to the potential for “external data [is used] to estimate counterfactual survival beyond the switching time-point” and suggests that “further research [to develop such methods] may be valuable” (TSD 24, April 2024). In this article, we proposed a new approach, the *ATSE*, which may indeed be prove valuable.

While RWE can be used as an external control arm to completely replace a crossover contaminated RCT control arm, this strategy discards a large amount of potentially valuable data (including the TTP times of those RCT subjects randomized to control) and can be susceptible to bias due to unmeasured confounding.³⁴ As illustrated in the simulation study, the *ATSE* approach leverages all the available data (and only the necessary external data) and consequently may be less impacted by confounding bias. This aligns with the current understanding that hybrid control arm studies (also known as “augmented RCTs”) should be considered a higher level of evidence than external control arm studies; see Gray et al.³⁵

If the assumptions underlying the treatment switching adjustment are suspect or the size of the external dataset is very large, then the ECA may be preferable to the *ATSE* approach. Alternatively, if the assumption of exchangeability for the subjects in the external data is suspect, then the TSE may be preferable to the *ATSE* approach. Another issue to consider is the possibility of immortal time bias and selection bias. To avoid immortal time bias and selection bias with the ECA approach, one must carefully select an appropriate “time-zero.” This can often be challenging (Jaksa et al.,²³ Fu et al.,³⁶ and Hernán et al.³⁷). With the proposed *ATSE* approach, selecting the appropriate “time-zero” should be relatively straightforward so long as the progression time is well defined. Figure 10 illustrates where the different methods might fit in terms of deciding upon the most appropriate approach.

When compared to the standard TSE approach, the *ATSE* approach has potential to obtain more precise estimation of survival outcomes particularly when sample sizes are small, and switching rates are high. However, the need for

strong assumptions remains. With TSE, the assumption of no unmeasured confounding implies that, conditional on all the observed covariates, a participant's decision to switch is independent of their PPS. With ATSE, an additional assumption is also required: Conditional on all observed covariates, individuals from the RCT and the external cohort must be exchangeable; see Bours.³⁸ The ATSE method also requires one to pre-specify a value for the decay factor, which will impact the overall amount of borrowing. Tan et al. recommend that this value be determined prior to analyzing the data by means of a simulation study, and based on Tan et al.'s simulation study, Sengupta et al.¹⁹ use a value of $c = 4$ for their case study. In our simulation study, we considered $c = 1, 4$, and 8 and found that the results did not differ substantially across the three different values. Following one's analysis, the amount of borrowing can be assessed using the "effective number of external events," see Sengupta et al.¹⁹


While the ATSE method represents a promising avenue for addressing treatment switching in clinical trials, its limitations must be acknowledged and carefully considered. One notable limitation is the assumption that treatment switching occurs exclusively at, or shortly after, disease progression. This assumption may not always hold true in real-world scenarios, where switching can occur for various reasons and at different points in time. Our simulation study was simply intended to demonstrate a proof of concept in simple scenarios, and as such, we did not investigate the impact of time-dependent confounding. The g-computation TSE method, proposed by Latimer et al.,³¹ offers a potential solution for situations where switching is not confined to the point of progression (or to another well-defined "secondary baseline"). More recently, Jackson et al. have proposed a variety of new approaches for TSE and suggest a method in which the secondary baseline is defined as the time of an individual's first subsequent treatment. Further research and simulation studies are needed to explore the feasibility of augmenting the g-computation TSE and other TSE approaches with external data. Our simulation study was also rather simplistic in that we simulated TTP times as being equal to (on average) about one-third of OS times. Future simulation studies could instead consider multi-state models (Jansen et al.³⁹).


Another potential limitation of ATSE might be the need for comprehensive data on all confounding variables. To avoid any bias due to unmeasured confounding, one must adjust for any factors that could simultaneously influence survival, switching, and participation in the RCT vs external cohorts. Identifying these factors could prove difficult, and the conservative approach of simply adjusting for all variables that could be considered important prognostic factors requires substantial data availability. The dynamic borrowing approach we propose within the ATSE method can help reduce the amount of bias due to unmeasured confounding, but cannot entirely eliminate it. Adjusting for TTP as a "proxy confounder" might help further mitigate the risk of bias due to unmeasured confounding. Finally, to calculate PPS for individuals in the external data, we emphasize that accurate recording of TTP is important and acknowledge that cancer progression is often poorly recorded in both hospital and registration data.^{40,41}


Acknowledgements

The research was performed without specific funding.

ORCID iDs

Harlan Campbell  <https://orcid.org/0000-0002-0959-1594>

Nicholas Latimer  <https://orcid.org/0000-0001-5304-5585>

Shannon Cope  <https://orcid.org/0000-0002-1721-9485>

Ethical approval and informed consent statements

There are no human participants in this article, and informed consent is not required.

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

Declaration of conflicting interests

The author(s) declared the following potential conflicts of interest with respect to the research, authorship, and/or publication of this article: Harlan Campbell, Jeroen P Jansen, and Shannon Cope are employees of Precision AQ.

Data availability statement

The data that support the findings of this study are simulated data, and the code used to simulate this data is available in the Appendix. Additional code is available from the corresponding author upon reasonable request.

Supplemental material

Supplemental material for this article is available online.

Note

- a. First, it was noted that there were no analyses done to understand the likelihood of unmeasured confounders (i.e. to determine if it was possible to fully adjust the external trial). Second, there were concerns that the external data were dated and possibly no longer relevant to the trial due to a long-term trend of improving survival. Finally, the adjusted treatment estimates seemed rather implausible (i.e. the results suggested that receiving treatment after disease progression had the same (if not higher) benefit on overall survival (OS) then when received before disease progression).

References

1. Newington L and Metcalfe A. Factors influencing recruitment to research: qualitative study of the experiences and perceptions of research teams. *BMC Med Res Methodol* 2014; **14**: 1–11.
2. Fojo T, Amiri-Kordestani L and Bates SE. Potential pitfalls of crossover and thoughts on iniparib in triple-negative breast cancer. *J Natl Cancer Inst* 2011; **103**: 1738–1740.
3. Chen EY and Prasad V. Crossover is not associated with faster trial accrual. *Ann Oncol* 2018; **29**: 776–777.
4. Henshall C, Latimer NR, Sansom L, et al. Treatment switching in cancer trials: issues and proposals. *Int J Technol Assess Health Care* 2016; **32**: 167–174.
5. Robins JM and Tsiatis AA. Correcting for non-compliance in randomized trials using rank preserving structural failure time models. *Commun Stat-Theory Methods* 1991; **20**: 2609–2631.
6. Robins JM and Finkelstein DM. Correcting for noncompliance and dependent censoring in an AIDS clinical trial with inverse probability of censoring weighted (IPCW) log-rank tests. *Biometrics* 2000; **56**: 779–788.
7. Latimer NR, Abrams KR, Lambert PC, et al. Adjusting for treatment switching in randomised controlled trials – a simulation study and a simplified two-stage method. *Stat Methods Med Res* 2017; **26**: 724–751.
8. Morden JP, Lambert PC, Latimer N, et al. Assessing methods for dealing with treatment switching in randomised controlled trials: a simulation study. *BMC Med Res Methodol* 2011; **11**: 1–20.
9. Watkins C, Huang X, Latimer N, et al. Adjusting overall survival for treatment switches: commonly used methods and practical application. *Pharm Stat* 2013; **12**: 348–357.
10. Latimer NR, Henshall C, Siebert U, et al. Treatment switching: statistical and decision-making challenges and approaches. *Int J Technol Assess Health Care* 2016; **32**: 160–166.
11. Morschhauser F, Ghesquieres H, Kamdar M, et al. Comparison of overall survival of lisocabtagene maraleucel (liso-cel) versus standard of care (SOC) adjusting for crossover in second-line (2L) R/R large B-cell lymphoma. *Hematol Oncol* 2023; **41**: 451–452.
12. Hawkins N, Muszbek N, Evans R, et al. Adjusting for treatment crossover in the MAVORIC trial: survival in advanced mycosis fungoides and Sézary syndrome. *J Comp Eff Res* 2022; **11**: 805–813.
13. Ishak KJ, Caro JJ, Drayson MT, et al. Adjusting for patient crossover in clinical trials using external data: a case study of lenalidomide for advanced multiple myeloma. *Value Health* 2011; **14**: 672–678.
14. Hoyle M, Rogers G, Garside R, et al. Commissioned by NHS R&D HTA Programme The clinical- and cost-effectiveness of lenalidomide for multiple myeloma in people who have received at least one prior therapy: an evidence review of the submission from Celgene. 2008. <https://researchers.mq.edu.au/en/publications/the-clinical-and-cost-effectiveness-of-lenalidomide-for-multiple->
15. Latimer NR and Abrams KR. NICE DSU technical support document 16: adjusting survival time estimates in the presence of treatment switching. 2014. <https://www.ncbi.nlm.nih.gov/books/NBK310374/>.
16. Struebing A, McKibbin C, Ruan H, et al. Augmenting external control arms using Bayesian borrowing: a case study in first-line non-small cell lung cancer. *J Comp Eff Res* 2024; **13**: e230175.
17. Mishra-Kalyani PS, Kordestani LA, Rivera DR, et al. External control arms in oncology: current use and future directions. *Ann Oncol* 2022; **33**: 376–383.
18. Li C, Ferro A, Mhatre SK, et al. Hybrid-control arm construction using historical trial data for an early-phase, randomized controlled trial in metastatic colorectal cancer. *Communications Medicine* 2022; **2**: 90.
19. Sengupta S, Ntambwe I, Tan K, et al. Emulating randomized controlled trials with hybrid control arms in oncology: a case study. *Clin Pharmacol Ther* 2023; **113**: 867–877.
20. Zhang J and Chen C. Correcting treatment effect for treatment switching in randomized oncology trials with a modified iterative parametric estimation method. *Stat Med* 2016; **35**: 3690–3703.
21. Latimer NR, White IR, Abrams KR, et al. Causal inference for long-term survival in randomised trials with treatment switching: should re-censoring be applied when estimating counterfactual survival times? *Stat Methods Med Res* 2019; **28**: 2475–2493.
22. Tennant PWG, Murray EJ, Arnold KF, et al. Use of directed acyclic graphs (DAGs) to identify confounders in applied health research: review and recommendations. *Int J Epidemiol* 2021; **50**: 620–632.
23. Jaksa A, Louder A, Maksymiuk C, et al. A comparison of 7 oncology external control arm case studies: critiques from regulatory and health technology assessment agencies. *Value Health* 2022; **25**: 1967–1976.
24. Ibrahim JG and Chen M-H. Power prior distributions for regression models. *Stat Sci* 2000; **15**: 46–60.
25. Hobbs BP, Sargent DJ and Carlin BP. Commensurate priors for incorporating historical information in clinical trials using general and generalized linear models. *Bayesian Analysis (Online)* 2012; **7**: 639.
26. Schmidli H, Gsteiger S, Roychoudhury S, et al. Robust meta-analytic-predictive priors in clinical trials with historical control information. *Biometrics* 2014; **70**: 1023–1032.

27. Alt EM, Chang X, Jiang X, et al. LEAP: the latent exchangeability prior for borrowing information from historical data. *Biometrics* 2024; **80**: ujae083.
28. Campbell H and Gustafson P. Discussion on “LEAP: the latent exchangeability prior for borrowing information from historical data” by Ethan M. Alt, Xiuya Chang, Xun Jiang, Qing Liu, May Mo, H. Amy Xia, and Joseph G. Ibrahim. *Biometrics* 2024; **80**: ujae084.
29. Tan WK, Segal BD, Curtis MD, et al. Augmenting control arms with real-world data for cancer trials: hybrid control arm methods and considerations. *Contemporary Clinical Trials Communications* 2022; **30**: 101000.
30. VanderWeele TJ. Principles of confounder selection. *Eur J Epidemiol* 2019; **34**: 211–219.
31. Latimer N, White I, Tilling K, et al. Improved two-stage estimation to adjust for treatment switching in randomised trials: g-estimation to address time-dependent confounding. *Stat Methods Med Res* 2020; **29**: 2900–2918.
32. Latimer N. NICE DSU Technical Support Document 14: Survival analysis for economic evaluations alongside clinical trials-extrapolation with patient-level data. 2011.
33. Greifer N and Greifer MN. Package ‘WeightIt’. *CRAN* <https://cranr-project.org/web/packages/WeightIt/index.html> 2019.
34. Burger HU, Gerlinger C, Harbron C, et al. The use of external controls: to what extent can it currently be recommended? *Pharm Stat* 2021; **20**: 1002–1016.
35. Gray CM, Grimson F, Layton D, et al. A framework for methodological choice and evidence assessment for studies using external comparators from real-world data. *Drug Saf* 2020; **43**: 623–633.
36. Fu EL, Harhay M, Schneeweiss S, et al. Starting right: aligning eligibility and treatment assignment at time zero when emulating a target trial. *Available at SSRN 5177135* 2025. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5177135.
37. Hernán MA, Sauer BC, Hernández-Díaz S, et al. Specifying a target trial prevents immortal time bias and other self-inflicted injuries in observational analyses. *J Clin Epidemiol* 2016; **79**: 70–75.
38. Bours MJL. A nontechnical explanation of the counterfactual definition of confounding. *J Clin Epidemiol* 2020; **121**: 91–100.
39. Jansen J, Incerti D and Trikalinos T. Multi-state network meta-analysis of cause-specific survival data. *Stat Med* 2023; **42**: 3371–3391.
40. Bright CJ, Lawton S, Benson S, et al. Data resource profile: the systemic anti-cancer therapy (SACT) dataset. *Int J Epidemiol* 2020; **49**: 15–15l.
41. Henson KE, Elliss-Brookes L, Coupland VH, et al. Data resource profile: national cancer registration dataset in England. *Int J Epidemiol* 2020; **49**: 16–16h.

Appendix

Augmented two-stage estimation for treatment switching in oncology trials: Leveraging external data for improved precision (Tables 3 to 6).

Table 3. Simulation study results for Scenarios 1–4 (no censoring; small external control, $N_{EC} = 200$).

Condition	Method	Percent bias in RMST			Empirical SE as a percentage of true RMST			RMSE as a percentage of true RMST		
		A	B	C	A	B	C	A	B	C
Scenario 1 Low treatment effect and moderate switching proportion, $N_{EC} = 200$	Oracle	0.30	0.30	0.30	5.70	5.70	5.70	5.70	5.70	5.70
	ITT	5.30	5.30	4.20	6.00	6.00	5.90	8.00	8.00	7.30
	TSE	0.50	0.50	2.10	7.60	7.60	7.20	7.60	7.60	7.50
	ECA	−0.30	5.00	−0.30	6.30	6.70	6.30	6.30	8.30	6.30
	ATSE ($c = 1$)	0.20	1.50	1.10	5.50	5.60	5.60	5.50	5.80	5.80
	ATSE ($c = 4$)	0.20	1.40	1.20	5.80	5.80	5.90	5.80	6.00	6.00
	ATSE ($c = 8$)	0.30	1.20	1.40	6.20	6.30	6.20	6.30	6.40	6.40
	ATSE ($c = 4$)	0.20	0.70	0.60	6.00	6.00	5.90	6.00	6.10	6.00
	adj. for TTP									
Scenario 2 High treatment effect and moderate switching proportion, $N_{EC} = 200$	Oracle	0.17	0.17	0.17	5.54	5.54	5.54	5.54	5.54	5.54
	ITT	14.24	14.24	11.28	6.38	6.38	6.13	15.61	15.61	12.84
	TSE	0.34	0.34	1.84	7.44	7.44	6.96	7.45	7.45	7.20
	ECA	−0.27	4.90	−0.27	6.27	6.60	6.27	6.28	8.22	6.28
	ATSE ($c = 1$)	0.18	1.52	1.03	5.44	5.49	5.49	5.44	5.69	5.58
	ATSE ($c = 4$)	0.18	1.31	1.11	5.63	5.67	5.66	5.63	5.82	5.76
	ATSE ($c = 8$)	0.25	1.07	1.28	6.07	6.10	5.96	6.07	6.19	6.09
	ATSE ($c = 4$)	0.24	0.73	0.60	5.93	5.96	5.83	5.93	6.01	5.86
	adj. for TTP									
Scenario 3 Low treatment effect and high switching proportion, $N_{EC} = 200$	Oracle	0.11	0.11	0.11	5.58	5.58	5.58	5.58	5.58	5.58
	ITT	7.28	7.28	6.18	6.01	6.01	5.93	9.44	9.44	8.56
	TSE	0.19	0.19	2.54	9.01	9.01	8.16	9.01	9.01	8.55
	ECA	−0.03	5.16	−0.03	6.24	6.54	6.24	6.24	8.33	6.24
	ATSE ($c = 1$)	0.14	2.12	1.29	5.54	5.62	5.66	5.54	6.01	5.80
	ATSE ($c = 4$)	0.08	1.74	1.45	5.81	5.91	5.93	5.81	6.16	6.10
	ATSE ($c = 8$)	0.17	1.34	1.83	6.66	6.79	6.58	6.67	6.92	6.83
	ATSE ($c = 4$)	0.25	0.96	0.70	6.28	6.33	6.23	6.28	6.40	6.27
	adj. for TTP									
Scenario 4 High treatment effect and high switching proportion, $N_{EC} = 200$	Oracle	0.11	0.11	0.11	5.66	5.66	5.66	5.66	5.66	5.66
	ITT	20.61	20.61	17.48	6.90	6.90	6.69	21.73	21.73	18.71
	TSE	0.10	0.10	2.48	9.08	9.08	8.12	9.08	9.08	8.48
	ECA	−0.14	5.07	−0.14	6.31	6.66	6.31	6.31	8.37	6.31
	ATSE ($c = 1$)	0.03	2.00	1.18	5.49	5.56	5.64	5.49	5.91	5.76
	ATSE ($c = 4$)	0.08	1.72	1.42	5.96	6.03	6.00	5.96	6.27	6.16
	ATSE ($c = 8$)	0.13	1.29	1.73	6.71	6.79	6.54	6.71	6.91	6.76
	ATSE ($c = 4$)	0.26	0.96	0.62	6.33	6.37	6.23	6.33	6.44	6.26
	adj. for TTP									

RMST: restricted mean survival time; SE: standard error; RMSE: root mean squared error; ITT: intention-to-treat; TSE: two-stage estimation; ECA: external control arm; ATSE: augmented two-stage estimation; TTP: time to progression.

Table 4. Simulation study results for Scenarios 5–8 (moderate censoring; small external control, $N_{EC} = 200$).

Condition	Method	Percent bias in RMST			Empirical SE as a percentage of true RMST			RMSE as a percentage of true RMST		
		A	B	C	A	B	C	A	B	C
Scenario 5 Low treatment effect and moderate switching proportion, $N_{EC} = 200$	Oracle	0.09	0.09	0.09	3.53	3.53	3.53	3.53	3.53	3.53
	ITT	2.56	2.56	2.15	3.53	3.53	3.53	4.35	4.35	4.13
	TSE	0.48	0.48	1.05	4.50	4.50	4.23	4.52	4.52	4.36
	ECA	0.08	2.59	0.08	3.69	3.66	3.69	3.69	4.48	3.69
	ATSE ($c = 1$)	0.47	0.98	0.75	3.36	3.37	3.38	3.39	3.51	3.47
	ATSE ($c = 4$)	0.46	0.91	0.78	3.47	3.46	3.48	3.50	3.58	3.57
	ATSE ($c = 8$)	0.51	0.84	0.87	3.68	3.68	3.65	3.71	3.78	3.75
	ATSE ($c = 4$) adj. for TTP	−0.02	0.19	0.20	3.72	3.72	3.70	3.72	3.72	3.70
Scenario 6 High treatment effect and moderate switching proportion, $N_{EC} = 200$	Oracle	−0.10	−0.10	−0.10	3.50	3.50	3.50	3.50	3.50	3.50
	ITT	5.52	5.52	4.61	3.45	3.45	3.46	6.51	6.51	5.76
	TSE	0.59	0.59	1.05	4.68	4.68	4.38	4.72	4.72	4.50
	ECA	−0.07	2.43	−0.07	3.59	3.56	3.59	3.59	4.31	3.59
	ATSE ($c = 1$)	0.62	1.14	0.77	3.38	3.37	3.43	3.43	3.56	3.52
	ATSE ($c = 4$)	0.62	1.07	0.81	3.48	3.47	3.50	3.53	3.63	3.59
	ATSE ($c = 8$)	0.63	1.00	0.86	3.72	3.71	3.68	3.78	3.84	3.78
	ATSE ($c = 4$) adj. for TTP	−0.43	−0.24	−0.25	3.91	3.90	3.86	3.94	3.90	3.87
Scenario 7 Low treatment effect and high switching proportion, $N_{EC} = 200$	Oracle	0.04	0.04	0.04	3.49	3.49	3.49	3.49	3.49	3.49
	ITT	3.29	3.29	2.89	3.46	3.46	3.46	4.77	4.77	4.51
	TSE	0.61	0.61	1.14	5.07	5.07	4.57	5.10	5.10	4.71
	ECA	−0.08	2.41	−0.08	3.59	3.57	3.59	3.59	4.31	3.59
	ATSE ($c = 1$)	0.56	1.24	0.78	3.25	3.25	3.30	3.30	3.48	3.39
	ATSE ($c = 4$)	0.60	1.20	0.86	3.41	3.41	3.44	3.46	3.62	3.54
	ATSE ($c = 8$)	0.61	1.05	0.94	3.71	3.73	3.67	3.76	3.87	3.79
	ATSE ($c = 4$) adj. for TTP	−0.07	0.20	0.08	3.94	3.95	3.84	3.94	3.95	3.84
Scenario 8 High treatment effect and high switching proportion, $N_{EC} = 200$	Oracle	−0.06	−0.06	−0.06	3.48	3.48	3.48	3.48	3.48	3.48
	ITT	7.53	7.53	6.61	3.38	3.38	3.40	8.25	8.25	7.44
	TSE	1.11	1.11	1.60	5.28	5.28	4.80	5.39	5.39	5.06
	ECA	0.00	2.51	0.00	3.62	3.61	3.62	3.61	4.40	3.61
	ATSE ($c = 1$)	0.99	1.71	1.13	3.42	3.41	3.50	3.56	3.81	3.68
	ATSE ($c = 4$)	1.00	1.63	1.21	3.56	3.55	3.62	3.70	3.91	3.81
	ATSE ($c = 8$)	1.09	1.56	1.38	3.95	3.95	3.91	4.09	4.25	4.14
	ATSE ($c = 4$) adj. for TTP	−0.52	−0.23	−0.30	4.04	4.03	4.00	4.07	4.03	4.01

RMST: restricted mean survival time; SE: standard error; RMSE: root mean squared error; ITT: intention-to-treat; TSE: two-stage estimation; ECA: external control arm; ATSE: augmented two-stage estimation; TTP: time to progression.

Table 5. Simulation study results for Scenarios 9–12 (no censoring; large external control, $N_{EC} = 1000$).

Condition	Method	Percent bias in RMST			Empirical SE as a percentage of true RMST			RMSE as a percentage of true RMST		
		A	B	C	A	B	C	A	B	C
Scenario 9 <i>Low treatment effect and moderate switching proportion, $N_{EC} = 1000$</i>	Oracle	−0.23	−0.23	−0.23	5.54	5.54	5.54	5.54	5.54	5.54
	ITT	4.83	4.83	3.77	5.80	5.80	5.71	7.54	7.54	6.84
	TSE	0.12	0.12	1.63	7.39	7.39	7.03	7.39	7.39	7.22
	ECA	0.04	5.23	0.04	2.79	2.93	2.79	2.79	6.00	2.79
	ATSE ($c = 1$)	−0.02	1.65	0.67	4.67	4.69	4.94	4.67	4.97	4.98
	ATSE ($c = 4$)	−0.01	1.59	0.70	4.77	4.79	5.03	4.77	5.04	5.07
	ATSE ($c = 8$)	0.01	1.46	0.76	4.94	4.99	5.17	4.94	5.20	5.23
	ATSE ($c = 4$) adj. for TTP	−0.06	0.61	0.18	5.50	5.50	5.53	5.50	5.54	5.53
Scenario 10 <i>High treatment effect and moderate switching proportion, $N_{EC} = 1000$</i>	Oracle	0.09	0.09	0.09	5.59	5.59	5.59	5.59	5.59	5.59
	ITT	14.68	14.68	11.58	6.45	6.45	6.22	16.04	16.04	13.14
	TSE	0.18	0.18	1.87	7.52	7.52	7.01	7.53	7.53	7.25
	ECA	0.02	5.23	0.02	2.86	3.02	2.86	2.86	6.04	2.86
	ATSE ($c = 1$)	0.04	1.72	0.82	4.71	4.72	4.95	4.71	5.02	5.02
	ATSE ($c = 4$)	0.05	1.65	0.85	4.81	4.82	5.03	4.81	5.09	5.10
	ATSE ($c = 8$)	0.06	1.51	0.91	5.03	5.08	5.20	5.03	5.29	5.28
	ATSE ($c = 4$) adj. for TTP	0.11	0.79	0.41	5.56	5.56	5.52	5.56	5.62	5.54
Scenario 11 <i>Low treatment effect and high switching proportion, $N_{EC} = 1000$</i>	Oracle	−0.07	−0.07	−0.07	5.70	5.70	5.70	5.70	5.70	5.70
	ITT	6.88	6.88	5.81	6.10	6.10	6.01	9.19	9.19	8.36
	TSE	−0.20	−0.20	2.07	9.03	9.03	8.11	9.03	9.03	8.37
	ECA	0.02	5.21	0.02	2.87	3.04	2.87	2.87	6.03	2.87
	ATSE ($c = 1$)	−0.06	2.32	0.75	4.23	4.26	4.56	4.23	4.85	4.62
	ATSE ($c = 4$)	−0.04	2.22	0.83	4.40	4.45	4.72	4.40	4.97	4.79
	ATSE ($c = 8$)	−0.07	1.92	0.91	4.83	4.95	5.08	4.83	5.31	5.16
	ATSE ($c = 4$) adj. for TTP	−0.01	0.93	0.31	5.71	5.73	5.74	5.71	5.80	5.75
Scenario 12 <i>High treatment effect and high switching proportion, $N_{EC} = 1000$</i>	Oracle	−0.13	−0.13	−0.13	5.70	5.70	5.70	5.70	5.70	5.70
	ITT	20.00	20.00	16.92	6.90	6.90	6.73	21.16	21.16	18.21
	TSE	−0.18	−0.18	2.18	9.00	9.00	8.06	9.00	9.00	8.35
	ECA	0.04	5.24	0.04	2.81	2.97	2.81	2.81	6.02	2.81
	ATSE ($c = 1$)	−0.03	2.35	0.79	4.25	4.28	4.56	4.25	4.89	4.63
	ATSE ($c = 4$)	0.03	2.29	0.89	4.41	4.46	4.70	4.41	5.01	4.78
	ATSE ($c = 8$)	0.03	2.03	1.05	4.89	5.02	5.00	4.89	5.42	5.11
	ATSE ($c = 4$) adj. for TTP	0.01	0.94	0.32	5.47	5.49	5.63	5.47	5.57	5.64

RMST: restricted mean survival time; SE: standard error; RMSE: root mean squared error; ITT: intention-to-treat; TSE: two-stage estimation; ECA: external control arm; ATSE: augmented two-stage estimation; TTP: time to progression.

Table 6. Simulation study results for Scenarios 13–16 (moderate censoring; large external control, $N_{EC} = 1000$).

Condition	Method	Percent bias in RMST			Empirical SE as a percentage of true RMST			RMSE as a percentage of true RMST		
		A	B	C	A	B	C	A	B	C
Scenario 13 <i>Low treatment effect and moderate switching proportion, $N_{EC} = 1000$</i>	Oracle	−0.05	−0.05	−0.05	3.51	3.51	3.51	3.51	3.51	3.51
	ITT	2.25	2.25	1.87	3.49	3.49	3.48	4.15	4.15	3.95
	TSE	0.27	0.27	0.79	4.57	4.57	4.22	4.58	4.58	4.29
	ECA	−0.05	2.46	−0.05	1.70	1.68	1.70	1.70	2.98	1.70
	ATSE ($c = 1$)	0.26	0.89	0.46	2.99	2.99	3.07	3.00	3.12	3.10
	ATSE ($c = 4$)	0.31	0.91	0.50	3.03	3.02	3.10	3.04	3.15	3.14
	ATSE ($c = 8$)	0.31	0.87	0.51	3.12	3.13	3.17	3.14	3.25	3.21
	ATSE ($c = 4$) adj. for TTP	−0.08	0.22	0.05	3.44	3.42	3.44	3.44	3.42	3.44
Scenario 14 <i>High treatment effect and moderate switching proportion, $N_{EC} = 1000$</i>	Oracle	−0.10	−0.10	−0.10	3.47	3.47	3.47	3.47	3.47	3.47
	ITT	5.82	5.82	4.84	3.44	3.44	3.43	6.76	6.76	5.93
	TSE	0.61	0.61	1.01	4.65	4.65	4.36	4.69	4.69	4.47
	ECA	0.00	2.50	0.00	1.64	1.63	1.64	1.64	2.99	1.64
	ATSE ($c = 1$)	0.57	1.25	0.68	3.07	3.03	3.13	3.12	3.28	3.21
	ATSE ($c = 4$)	0.55	1.22	0.68	3.10	3.08	3.19	3.15	3.31	3.26
	ATSE ($c = 8$)	0.56	1.18	0.71	3.18	3.16	3.25	3.22	3.38	3.33
	ATSE ($c = 4$) adj. for TTP	−0.62	−0.30	−0.43	3.59	3.56	3.63	3.64	3.57	3.65
Scenario 15 <i>Low treatment effect and high switching proportion, $N_{EC} = 1000$</i>	Oracle	−0.04	−0.04	−0.04	3.49	3.49	3.49	3.49	3.49	3.49
	ITT	2.99	2.99	2.61	3.44	3.44	3.45	4.56	4.56	4.32
	TSE	0.52	0.52	1.08	5.41	5.41	4.62	5.44	5.44	4.74
	ECA	0.00	2.50	0.00	1.63	1.61	1.63	1.63	2.97	1.63
	ATSE ($c = 1$)	0.49	1.32	0.64	2.76	2.75	2.85	2.81	3.05	2.92
	ATSE ($c = 4$)	0.51	1.31	0.67	2.83	2.83	2.90	2.88	3.12	2.98
	ATSE ($c = 8$)	0.52	1.25	0.71	3.05	3.05	3.04	3.09	3.30	3.12
	ATSE ($c = 4$) adj. for TTP	−0.10	0.29	0.01	3.38	3.38	3.36	3.38	3.39	3.36
Scenario 16 <i>High treatment effect and high switching proportion, $N_{EC} = 1000$</i>	Oracle	0.00	0.00	0.00	3.51	3.51	3.51	3.51	3.51	3.51
	ITT	7.59	7.59	6.66	3.43	3.43	3.46	8.32	8.32	7.50
	TSE	1.22	1.22	1.63	5.35	5.35	4.78	5.48	5.48	5.04
	ECA	0.00	2.51	0.00	1.64	1.63	1.64	1.64	2.99	1.64
	ATSE ($c = 1$)	0.97	1.83	1.03	2.82	2.75	2.97	2.98	3.30	3.14
	ATSE ($c = 4$)	0.93	1.77	1.02	2.86	2.83	3.05	3.01	3.34	3.22
	ATSE ($c = 8$)	1.00	1.77	1.11	3.07	3.04	3.19	3.23	3.52	3.38
	ATSE ($c = 4$) adj. for TTP	−0.46	−0.05	−0.34	3.55	3.53	3.57	3.58	3.53	3.58

RMST: restricted mean survival time; SE: standard error; RMSE: root mean squared error; ITT: intention-to-treat; TSE: two-stage estimation; ECA: external control arm; ATSE: augmented two-stage estimation; TTP: time to progression.