# Adjusting Survival Time Estimates to Account for Treatment Switching in Randomized Controlled Trials—an Economic Evaluation Context: Methods, Limitations, and Recommendations

*Nicholas R. Latimer, PhD, Keith R. Abrams, PhD, Paul C. Lambert, PhD, Michael J. Crowther, MSc, Allan J. Wailoo, PhD, James P. Morden, MSc, Ron L. Akehurst, MSc, Michael J. Campbell, PhD*

***Background.*** *Treatment switching commonly occurs in clinical trials of novel interventions in the advanced or metastatic cancer setting. However, methods to adjust for switching have been used inconsistently and potentially inappropriately in health technology assessments (HTAs).* ***Objective.*** *We present recommendations on the use of methods to adjust survival estimates in the presence of treatment switching in the context of economic evaluations.* ***Methods.*** *We provide background on the treatment switching issue and summarize methods used to adjust for it in HTAs. We discuss the assumptions and limitations associated with adjustment methods and draw on results of a simulation study to make recommendations on their use.* ***Results.*** *We demonstrate that methods used to adjust for treatment switching have important limitations and often produce bias in realistic scenarios. We present an analysis framework that aims to increase the probability that suitable adjustment methods can be identified on a case-by-case basis. We recommend that the characteristics of clinical trials, and the treatment switching mechanism observed within them, should be considered alongside the key assumptions of the adjustment methods. Key assumptions include the ''no unmeasured confounders'' assumption associated with the inverse probability of censoring weights (IPCW) method and the ''common treatment effect'' assumption associated with the rank preserving structural failure time model (RPSFTM).* ***Conclusions.*** *The limitations associated with switching adjustment methods such as the RPSFTM and IPCW mean that they are appropriate in different scenarios. In some scenarios, both methods may be prone to bias; ''2-stage'' methods should be considered, and intention-to-treat analyses may sometimes produce the least bias. The data requirements of adjustment methods also have important implications for clinical trialists.*
***Key words:*** *survival analysis; prediction; treatment switching; treatment crossover; economic evaluation; modeling; technology assessment; statistical methods.* ***(Med Decis Making XXXX;XX:XXX–XXX)***

Treatment switching, in which patients randomized to the control group of a clinical trial are permitted to switch to the experimental treatment group, is common in trials of oncology treatments for both ethical and practical reasons. Ethically, when there are no other nonpalliative treatments available, it may be deemed inappropriate to deny control group patients the new treatment if interim analyses indicate a positive treatment effect. Practically, it may be difficult to recruit patients to a trial that does not allow treatment switching. In addition, pharmaceutical companies have responded to incentives associated with the acceptance of progression-free survival (PFS) as a primary end point for drug regulatory approval by agencies such as the US Food and Drug Administration (FDA) and the European Medicines Agency (EMA).[1,2] Randomized controlled trials (RCTs) of cancer treatments are often powered to investigate the difference in PFS rather than overall survival (OS). Hence, there is less motivation for pharmaceutical companies to ensure


DOI: 10.1177/0272989X13520192

that randomized groups are maintained beyond disease progression. Treatment switching is likely to cause more serious problems for health technology assessment (HTA) agencies than for licensing bodies. While showing an OS advantage may not be essential for gaining a license, a lifetime horizon is generally advocated in economic evaluations, especially for interventions that impact survival[3–6]; thus, estimates of the treatment effect on OS are required.

Typically, evidence on the effectiveness of new treatments is taken from RCTs and used within economic models to generate cost-effectiveness estimates. However, when patients in the control group switch to, and benefit from, the experimental treatment, a standard intention-to-treat (ITT) analysis (a comparison of groups as randomized) will underestimate the ''true'' survival benefit associated with the new treatment, that is, the benefit that would have been observed if switching had not been permitted. Hence, such an analysis would not meet the requirements of the economic evaluation decision problem, whereby a state of the world in which the new therapy is used for treatment is compared to one in which it is not. Statistical methods have been developed specifically to take into account treatment switching, but these have been little used in HTAs, as demonstrated in Appendix A. In this article, we summarize available treatment switching adjustment methods and assess their assumptions, limitations, and practical applicability specifically for an economic evaluation context. Our objective is to make recommendations on the use of switching adjustment methods.

In this article, treatment switching is defined as the switch from a control treatment to an experimental treatment by patients randomized to the control group of an RCT. It is worthy of note that some authors use the term ''treatment crossover'' rather than ''treatment switching''; here, we have used ''switching'' because ''crossover'' may evoke crossover trials, which are a different entity. As defined here, treatment switching does not involve experimental group patients switching to the control treatment or patients randomized to either group receiving other poststudy treatments. The reason that these treatment changes are not included within our definition of treatment switching is that they can both form part of a realistic treatment pathway, meaning that an appraisal of the relevant economic evaluation decision problem is still possible. Generally, an economic evaluation seeks to compare a state of the world in which the novel intervention is used and is given to a cohort of indicated patients to a state of the world in which the novel intervention is not used and standard treatments are received. If an experimental group patient discontinues the novel therapy and receives a standard treatment (either that received in the control group or a separate standard treatment), this is likely to have occurred because of treatment failure, toxicity, tolerability, or adverse events. Such events and subsequent treatment switches are likely to occur in reality, and therefore, they form a relevant part of the analysis of outcomes in the state of the world in which the new treatment is available. Hence, in general, we would not wish to adjust for these treatment changes in our economic analysis. Similarly, if control (or experimental) group patients received poststudy therapies that do not include the experimental treatment, this reflects a realistic treatment pathway, and we would not wish to adjust for this in our economic analysis. Even if differential proportions of patients receive different poststudy therapies, this may reflect appropriate treatment pathways given the initial treatment.

Received 7 March 2013 from the School of Health and Related Research, University of Sheffield, Sheffield, UK (NRL, AJW, RLA, MJC); Department of Health Sciences, University of Leicester, Leicester, UK (KRA, PCL, MJC); Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden (PCL); and Clinical Trials and Statistics Unit (ICR-CTSU), Division of Clinical Studies, The Institute of Cancer Research, London, UK (JPM). Parts of this work have been presented at the following conferences: International Society for Pharmacoeconomics and Outcomes Research, 15th Annual European Conference, Berlin, Germany, November 2012; and Health Technology Assessment International, 9th Annual Meeting, Bilbao, Spain, June 2012. Nicholas Latimer was funded by a National Institute for Health Research (NIHR) Doctoral Research Fellowship (DRF 2009-02-82). This article presents independent research funded by the NIHR. The views expressed are those of the authors and not necessarily those of the National Health Service, the NIHR, or the Department of Health. Keith Abrams is partially supported as a senior investigator by the NIHR (NI-SI-0508-10061). Michael Crowther is funded by an NIHR Doctoral Research Fellowship (DRF-2012-05-409). The Pharmaceutical Oncology Initiative, a group of pharmaceutical companies that are part of the Association of the British Pharmaceutical Industry, provided financial support for the simulation study referred to in this article. Financial support for this study was provided in part by grants from the NIHR and the Pharmaceutical Oncology Initiative. The funding agreement ensured the authors' independence in designing the study, interpreting the data, writing, and publishing the report. James Morden works for the ICR-CTSU, which receives core funding from Cancer Research UK. Revision accepted for publication 18 December 2013.

Supplementary material for this article is available on the *Medical Decision Making* website at http://mdm.sagepub.com/supplemental.

Address correspondence to Nicholas R. Latimer, PhD, School of Health and Related Research, University of Sheffield, Regent Court, 30 Regent Street, Sheffield, S1 4DA, UK; telephone: + 44 (0) 114 222 0821; e-mail: n.latimer@shef.ac.uk.

Unless this can be shown not to be the case, it would be inappropriate to adjust for these differences in the economic analysis.

In addition, in this article, we focus on survival or time-to-event outcomes. We recognize that these are not the only important outcomes that may be affected by treatment switching. For instance, costs and quality of life scores collected within an RCT and attributed to randomized groups will be subject to confounding. Adjusting for these outcomes lies outside the scope of this article, although we refer to this again in our discussion.

First, we set out the treatment switching problem in the context of an economic evaluation. Then, we summarize the key switching adjustment methods, outlining their pivotal assumptions and limitations as well as their practical applicability in an economic evaluation context. We include details on the use of these methods in an HTA context based on a review of technology appraisals (TAs) undertaken in the UK by the National Institute for Health and Care Excellence (NICE). This review is presented in Appendix A. Results of a previously reported simulation study are then drawn upon to make recommendations on the practical use of adjustment methods in the form of an analysis framework before final conclusions are made.

Research assessing the performance of switching adjustment methods has previously been published in the statistical literature,[7] but the use of these methods to inform economic models represents a neglected research area. Different adjustment methods provide different statistical outputs, which impact the ways in which these can be incorporated within an economic model. Our research fills this gap by considering adjustment methods specifically in the context of an economic evaluation.

## TREATMENT SWITCHING: THE PROBLEM

Treatment switching is an important problem for economists and decision makers because it typically leads to a treatment pathway that is not relevant for the decision problem defined in an HTA. Treatment switching causes a mismatch between what has been studied in the clinical trial and the economic analysts' decision problem; the comparator arm becomes contaminated due to treatment switching.

In this article, we define bias as the difference (error) between the estimated treatment effect and the effect that would have been observed in the absence of treatment switching. The bias that may be created by treatment switching and the theoretical problems that it creates for the economic analysis are illustrated in Figure 1. The first 2 rows (''Control Treatment'' and ''Intervention'') illustrate the ''perfect'' trial in which no treatment switching occurs. Survival time is on the x-axis, and in this example, the new intervention extends PFS and postprogression survival (PPS). This results in the ''true OS difference'' identified in the diagram. In this case, a standard ITT analysis will usually give us the information that we need for our economic model (ignoring any need for extrapolation), as this perfectly satisfies the economic evaluation decision problem of comparing a state of the world in which the experimental treatment is available to one in which the experimental treatment is not available. However, the third row (''Control ⟶ Intervention'') demonstrates what may happen to survival in the control group if treatment switching is permitted (in this case, after disease progression). PPS is extended compared to the ''Control Treatment'' comparator, under the assumption that some control group patients switch and benefit from the new intervention after disease progression. The result of this is that the OS difference observed in the RCT's ITT analysis (labeled ''RCT OS difference'' in Figure 1) is smaller than the true OS difference that would have been observed if no treatment switching had occurred, and the ITT analysis would not appropriately address the economic evaluation decision problem. The simple ITT analysis will result in bias equal to the difference between the ''true OS difference'' and the ''RCT OS difference'' when treatment switching occurs. The extent of this bias will be unknown, as the true OS difference will be unobserved. However, it is clear that provided that switching patients benefit to any extent from the new intervention, some bias will exist. An economic evaluation that relied on this ITT analysis would produce inaccurate cost-effectiveness results (in this case, the incremental cost-effectiveness ratio [ICER] would be overestimated), and inappropriate resource allocation decisions may be made. An economic evaluation that incorporated the costs of the treatment to which the patient switched may dilute this bias, but the extent to which this would reflect an accurate estimate of what the ICER would have been in the absence of switching would be unknown because of the likelihood that switchers are selected based on prognosis and their potentially reduced capacity to benefit: these issues may cause the ICER to be importantly different in switchers compared to patients randomized to the experimental group. To address the economic evaluation decision problem, it would be preferable to accurately adjust survival estimates for switching and to exclude the costs of switching treatments.

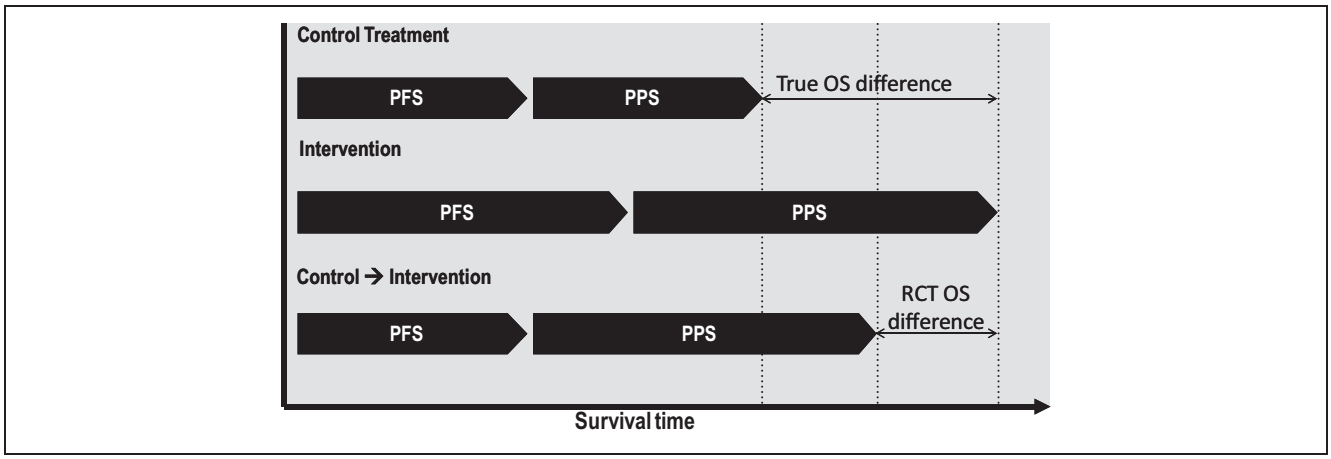


*Figure 1 The potential impact of treatment switching illustrated. OS = overall survival; PFS = progression-free survival; PPS = postprogression survival; RCT = randomized controlled trial.*

## SUMMARY OF METHODS

In this section, we introduce various treatment switching adjustment methods. We begin with relatively simple methods before moving on to more complex methods. The simpler methods are more commonly used in HTAs, as demonstrated by our review of NICE TAs presented in Appendix A. We reviewed all NICE TAs that had been completed by December 2009 and identified those that were in the area of advanced or metastatic cancer. Forty-five TAs were identified and included in the review, and of these, treatment switching occurred in 25 (55.6%). The methods used to address treatment switching in these appraisals are summarized in Table 1. Not all the simple methods listed in Table 1 are described in detail in this section because our focus is on methods that attempt to adjust observed survival estimates to account for switching rather than methods that do not make any use of the observed postdisease progression data (such as those that only model PFS, assume equal OS or an equal risk of death following disease progression, or simply include the costs of the treatment to which the patient switched in the economic model).

Here, we discuss the key assumptions and limitations of the key methods before considering how they may be incorporated within an economic model. We focus on the key principles of the methods rather than their mathematics, although further details on the more complex methods are provided in Appendix B.

### Simple Methods

*Intention to treat.* An ITT analysis does not attempt to adjust for treatment switching but was used in 7 (28%) of the 25 NICE TAs that were affected by treatment switching (Table 1 and Appendix A). Groups are compared as randomized, and thus, the randomization balance of the trial is respected. The ITT analysis represents a valid comparison of randomized groups, but in the presence of treatment switching, this is unlikely to be what is required for an economic evaluation because the ''true'' survival benefit associated with the novel intervention will be diluted due to the switching of control group patients to the novel therapy.

*Per protocol: excluding and censoring switchers.* In 11 (44%) of the 25 NICE TAs reviewed, a per-protocol analysis was used to adjust for treatment switching (Table 1 and Appendix A). Data from patients who switched were either excluded entirely from the analysis or were censored at the point of the switch. Such analyses are prone to selection bias because the randomization balance between groups is broken if switching is associated with prognostic patient characteristics, for instance, if patients with either a good or poor prognosis are more likely to switch.[8,9] This is highly likely in the case of treatment switching in clinical trials; clinicians decide whether it is appropriate for individual patients to switch, and this decision will be made based on patient characteristics rather than being random.

### Complex Methods

*Inverse probability of censoring weights (IPCW).* The IPCW method has been used in recent HTAs.[10,11] It represents an approach for adjusting estimates of a treatment effect in the presence of

**Table 1** Methods Used to Account for Switching in NICE TAs (2000–2009)

| Method | No. of TAs that Use the Method |
|---|---|
| ''Simple'' methods | |
| Intention-to-treat analysis (no attempt to adjust for switching) | 7 (TAs 3, 30, 55, 91, 124, 162, 172) |
| Censored patients | 6 (TAs 28, 86, 129, 169, 178, 179) |
| Excluded patients | 5 (TAs 34, 70, 86, 169, 178) |
| Included costs of switching treatments | 4 (TAs 101, 116, 118, 121) |
| Modeled based on progression-free survival, not overall survival | 2 (TAs 6, 33) |
| Used sequencing models | 2 (TAs 93, 176) |
| Applied same risk of death on disease progression | 1 (TA 118) |
| Assumed equal overall survival for the 2 treatment groups | 1 (TA 119) |
| More complex methods | |
| Rank preserving structural failure time model | 1 (TA 179) |
| Adjusted survival estimates using a case-mix approach | 1 (TA 34) |
| Used external data | 1 (TA 171) |

Note: For a discussion of the use of these methods in National Institute for Health and Care Excellence (NICE) technology appraisals (TAs) and a commentary on the use of complex methods in more recent appraisals, see Appendix A. Also note that the numbers in this table do not sum to 25 because in 6 TAs, more than 1 method was used.

any type of informative censoring. In the context of treatment switching, patients are artificially censored at the time of switching, and remaining observations are weighted based on covariate values and a model of the probability of being censored. This allows patients who have not been artificially censored to be weighted in order to reflect their similarities to patients who have been censored in an attempt to remove selection bias.

The key assumption made by the IPCW method is the ''no unmeasured confounders'' assumption; that is, data must be available on all baseline and time-dependent prognostic factors for mortality that independently predict informative censoring (switching), and models of censoring risk must be correctly specified.[12] In practice, this is unlikely to be perfectly true, but the method is likely to work adequately if the ''no unmeasured confounders'' assumption is approximately true; that is, there are no important independent predictors missing. If this is the case, the selection bias associated with the dependence between censoring and failure can be corrected by replacing the Kaplan-Meier estimator, log-rank test, and Cox partial likelihood estimator of the hazard ratio (HR) with their IPCW versions.[12]

The ''no unmeasured confounders'' assumption represents a key limitation of the IPCW method. It cannot be tested using the observed data[13,14] and is particularly problematic in an RCT context. The IPCW method represents a type of marginal structural model, which was originally developed for use with observational data.[15,16] Typically, RCT datasets are much smaller than observational datasets, and when fewer data are available (particularly on control group patients who do not switch), the IPCW method may become less stable, and confidence intervals may become wide. In addition, some key predictors of treatment switching are usually not collected in RCTs (such as patient preference for switching), and often, data collection on key indicators is stopped at some point (e.g., upon treatment discontinuation or disease progression), which hampers the applicability of the IPCW method. Finally, the IPCW method cannot work if there are levels of any covariates that ensure (i.e., the probability equals 1) that treatment switching will occur.[14–16]

*Rank preserving structural failure time model (RPSFTM).* The RPSFTM method was designed specifically for an RCT context and has been used recently in HTAs.[10,11,17] It uses a counterfactual framework to estimate the causal effect of the treatment in question,[18] where counterfactual survival times refer to those that would have been observed if no treatment had been given. It is assumed that counterfactual survival times are independent of treatment group, and g-estimation is used to determine a value for the treatment effect that satisfies this constraint. The RPSFTM is an instrumental variables (IV) method; such methods are often used when the data available are unlikely to capture all factors that predict both treatment and outcome (i.e., the ignorability assumption does not hold). In the context of treatment switching, where switching is highly likely to be associated with prognostic factors, this is likely to be the case. Under the IV

approach, the instrument is a variable (in this case, the randomized treatment group) that is predictive of the treatment that is used to estimate causal treatment effects. The instrument must affect the outcome only through its effect on an intermediate variable (here, treatment), which is an assumption that is known as the exclusion restriction (see Hernan and Robins[19] for further discussion on IV methods).

The RPSFTM does not rely on the ''no unmeasured confounders'' assumption and identifies the treatment effect using only randomization of the trial, observed survival, and observed treatment history. It is assumed that if 2 patients have the same observed event time and neither has received treatment, those 2 patients would also have the same event time if they both received treatment. This assumption is linked to the associated assumptions that the treatment effect (an ''acceleration factor,'' or ''time ratio'') is equal (relative to the time for which the treatment is taken) for all patients no matter when the treatment is received (the ''common treatment effect'' assumption) and that randomization of the trial means that there is only random variation between treatment groups at baseline, apart from the treatment allocated; untreated survival times must be independent of the randomized treatment group.[18] The RPSFTM's primary limitations involve the ''common treatment effect'' assumption and the randomization assumption. The latter should be reasonable in the context of an RCT, but the potential remains for important differences at baseline in small and larger trials.[20] It is therefore relevant to note that it is possible to adjust for baseline covariates within an RPSFTM analysis, which is useful to increase power.[21] The ''common treatment effect'' assumption is more problematic. If patients who switch to the experimental treatment partway through the trial receive a different treatment effect compared to patients originally randomized to the experimental group, the RPSFTM estimate of the treatment effect received by patients in the experimental group will be biased. Given that treatment switching is often only permitted after disease progression, at which time the capacity for a patient to benefit may be different compared to before progression, the ''common treatment effect'' assumption may not be clinically plausible. As for the ''no unmeasured confounders'' assumption, it is unlikely that the ''common treatment effect'' assumption will ever be exactly true. However, of more concern is whether the assumption is likely to be approximately true, that is, that the treatment effect received by switchers can at least be expected to be similar to the effect received by patients initially randomized to the experimental group.

*Iterative parameter estimation (IPE).* Branson and Whitehead[22] extended the RPSFTM method using parametric methods, developing a novel IPE procedure. The same accelerated failure time model is used, but a parametric failure time model is fitted to the original unadjusted ITT data to obtain an initial estimate of the treatment effect. The failure times of switching patients are then re-estimated using this, and this iterative procedure continues until the new estimate is very close to the previous estimate, at which point the process is said to have converged.[22] To our knowledge, this method has not yet been used in a published HTA.

The IPE procedure makes similar assumptions as the RPSFTM method; for example, the randomization assumption is made, as is the ''common treatment effect'' assumption. An additional assumption is that survival times follow a parametric distribution, and thus, it is important to identify suitable parametric models, which in itself can be problematic.[23]

## Alternative ''2-Stage'' Methods

In addition to the ''standard'' adjustment methods described so far, ''2-stage'' methods might be considered; to our knowledge, these have not yet been used in HTAs. These methods involve first estimating a treatment effect specific to switching patients and then using this to derive a counterfactual dataset unaffected by switching. Then, a treatment effect specific to patients randomized to the experimental group can be estimated. Robins and Greenland[13] and Yamaguchi and Ohashi[14] have previously used such an approach, making use of a structural nested failure time model (SNM) with g-estimation to estimate the treatment effect in switchers. The SNM is essentially an observational version of the RPSFTM and attempts to account for time-dependent confounding using the ''no unmeasured confounders'' assumption. It therefore has similar limitations as the IPCW.

A previously unused 2-stage approach that does not rely on g-estimation may provide a good fit to the treatment switching mechanism often observed in oncology RCTs. When switching is only permitted after disease progression, the time of progression can be used as a secondary baseline. Using this secondary baseline, a parametric accelerated failure time model (such as a Weibull model) that includes covariates measured at the time of progression could be fitted

to the postprogression control group data. This model could be used to estimate the effect of switching to the treatment after progression by contrasting outcomes in those control group patients who switch after progression with those who do not. The resulting acceleration factor can then be used to ''shrink'' the survival times of switching patients in order to derive a counterfactual dataset unaffected by switching. These methods effectively recognize that the clinical trial is randomized up until the point of disease progression, but beyond that point, it essentially becomes an observational study. This is a simplification of the method used by Robins and Greenland[13] and Yamaguchi and Ohashi[14] because no attempt is made to adjust for time-dependent confounding beyond disease progression. However, if switching is likely to happen soon after disease progression, any time-dependent confounding associated with the lag between disease progression and treatment switch would be small. Such a method may not be generalizable because it is reliant on the ability to identify a secondary baseline, and it requires the ''no unmeasured confounders'' assumption to hold at the point of the secondary baseline. However, it does not require data to be collected on these confounders at other time points and does not make the ''common treatment effect'' assumption.

### Summary

It is clear that alternative complex adjustment methods make very different assumptions and work in very different ways; hence, they are likely to produce different results. This has been demonstrated in HTAs; in the NICE appraisal of pazopanib for the first-line treatment of metastatic renal cell carcinoma (RCC), the IPCW method produced an ICER of approximately £49,000 per quality-adjusted life year (QALY) gained, whereas the RPSFTM method produced an ICER of approximately £33,000 per QALY gained (see Appendix A for further details).[10] Our review of NICE appraisals presented in Appendix A demonstrates that there has been a trend towards using more complex methods in HTAs, but there remains evidence of uncertainty around which methods are appropriate for adjusting to treatment switching as well as an important lack of understanding of what these methods entail. For example, in the NICE appraisals of pazopanib for the first-line treatment of metastatic RCC and of everolimus for the second-line treatment of advanced RCC, the weakness of the IPCW method due to its ''no unmeasured confounders'' assumption was highlighted, whereas the ''common treatment effect'' assumption made by the RPSFTM method was not discussed in any detail.[10,11] Hence, while the RPSFTM method appeared to be preferred in these appraisals, the advantages and disadvantages associated with each method did not appear to have been fully taken into account, and it is not clear that the most appropriate switching adjustment method was identified. Two-stage methods appear to be potentially useful methods that have not previously been used in HTAs.

## APPLICATION TO ECONOMIC EVALUATIONS

### Theoretical Limitations

It is important to consider the theoretical limitations associated with the treatment switching adjustment methods when considering their suitability for use within an economic evaluation. For the IPCW and 2-stage methods, this involves a consideration of the plausibility of the ''no unmeasured confounders'' assumption. Although this assumption cannot be tested, an assessment of the measured covariates alongside findings from previous studies in similar disease areas, combined with an elicitation of expert clinical opinion, may provide valuable information. The treatment switching mechanism within the trial of interest should also be explored to ascertain how and why treatment switching decisions were made, as this may provide information on whether data on key switching indicators were collected. Linked to this data issue is that of sample size and event numbers. The IPCW method bases its adjustment on the survival experiences of control group patients who do not switch treatments; if almost all patients switch, and/or very few events are observed in patients who do not switch, the method is unlikely to perform reliably.

For the RPSFTM and IPE methods, the clinical and biological plausibility of the ''common treatment effect'' assumption is critical. In circumstances where treatment switching occurs after disease progression, it may not be credible to assume that switchers, who now suffer a more advanced disease, receive the same benefit (per unit of time) from treatment as those in the experimental group who received the treatment from randomization. In an attempt to relax the ''common treatment effect'' assumption, analysts have attempted to apply a multiparameter version of the RPSFTM. However, these have not been successful, with meaningful point estimates for causal effects difficult to determine.[13,24,25] While

some assessment of the ''common treatment effect'' assumption may be made using trial data (e.g., by estimating the treatment effect received by switchers compared to nonswitchers), such analyses are likely to be prone to time-dependent confounding and are therefore unreliable. If patients with varying levels of disease progression were randomized to the trial of interest, comparing the treatment effect in groups based on the initial disease stage may be useful, although in end-stage metastatic cancer trials, this may not be possible. Hence, understanding the mechanism of action of the intervention and eliciting clinical expert opinion on its likely effectiveness at different points of the disease progression pathway are important.

Use of the RPSFTM and IPE methods is also problematic if the comparator treatment used in the RCT is active. The RPSFTM and IPE counterfactual survival model requires that patients are either ''on treatment'' or ''off treatment'' at any 1 time. If patients in the control group receive an active treatment followed by supportive care upon treatment failure, the ''off treatment'' category represents more than 1 type of treatment, and the counterfactual survival model is not appropriate unless additional causal parameters are added to the model, but as stated above, attempts to apply multiparameter RPSFTMs have not been successful. Standard RPSFTM or IPE methods could still be applied, but several important assumptions about treatment strategies and their effectiveness in the experimental and control groups would be required. Linked to this, the RPSFTM and IPE counterfactual survival model assumes that the treatment effect is only received while a patient is ''on treatment''; it disappears as soon as treatment is discontinued. The clinical plausibility of this assumption should be considered. If a continuing treatment effect is expected, the RPSFTM or IPE methods could be applied, assuming a lagged treatment effect or on a ''treatment group'' basis in which patients in the experimental group are always considered to be ''on treatment'' and patients who switch remain ''on treatment'' from the time of switching until death. This analysis ignores treatment discontinuation times and requires there to be a common treatment effect associated with the sequence of treatments received by patients randomized to the experimental group and the sequence of treatments received by switchers after the point of switching. Any benefits associated with poststudy treatments will be attributed to the experimental treatment, although similarly, any benefits from poststudy treatments received by control group nonswitchers would be attributed to the control group. If the poststudy treatments received in all groups represent realistic treatment pathways, this approach may appropriately address the economic evaluation decision problem particularly if the costs of the poststudy treatments are also incorporated within the economic model. Hence, such an approach might be considered if the comparator is active or if a continuing treatment effect is expected.

It is worthy of note that the randomization-based methods (RPSFTM and IPE) typically lose power in the presence of treatment switching, like the ITT analysis. By design, they maintain the significance level associated with the ITT analysis, and therefore, their confidence intervals are often relatively wide. Observation-based methods such as the IPCW and 2-stage methods are not restricted in this way, but their confidence intervals may also be wide if data are relatively sparse.

### Practical Limitations

The practical limitations associated with combining treatment switching adjustment methods with economic evaluations must also be considered. Latimer[23] provided recommendations on how the extrapolation of survival data should be undertaken for use in economic models. Two main approaches were described: extrapolation using parametric models fitted independently to treatment groups; and extrapolation undertaken based on a proportional treatment effect assumption, whereby 1 parametric model is fitted to both treatment groups combined, with the treatment group included as a covariate. Issues with both of these arise when treatment switching adjustment methods are used. The RPSFTM, IPE, and 2-stage methods provide a counterfactual dataset that is adjusted for treatment switching, and thus, either extrapolation approach can be undertaken. However, White and others[24] demonstrated that recensoring is required for the RPSFTM and IPE methods to avoid bias, and this is also true for 2-stage methods. Recensoring is required because a positive or negative treatment effect may increase or decrease the probability that the survival time of an individual is censored, and where treatment switching occurs, the treatment received is likely to be associated with the prognosis. This means that counterfactual censoring times may be related to the prognosis and may therefore be informative (see Appendix B for more details).[24] Recensoring involves data being recensored at an earlier time point to avoid informative censoring and is therefore associated with a loss of longer term survival information.

Some observed events will become censored if the recensoring time is shorter than the counterfactual event time. The time point at which recensoring occurs is related to the magnitude of the estimated treatment effect; the larger the treatment effect, the earlier the recensoring time point. Loss of long-term information is likely to be detrimental to the extrapolation of survival data, which is of particular importance in the context of economic evaluations because of the requirement to estimate the mean survival advantages associated with novel interventions.[3–6,23,26] In addition, recensoring may lead to biased estimates of the ''average'' treatment effect in circumstances where proportional treatment effect assumptions do not hold because longer term data on the effect of treatment may be lost.

The IPCW method provides an estimate of the treatment effect in the form of an adjusted HR as well as a weighted Kaplan-Meier (WKM) curve, which is associated with a counterfactual dataset. However, it is not simple to fit parametric models to the IPCW counterfactual dataset because of the weightings associated with each observation. Novel methods for the extraction of survival times from Kaplan-Meier curves could be used to generate a replacement counterfactual dataset using the WKM curve,[27] after which any of the extrapolation methods described by Latimer[23] could be applied. Alternatively, a variation on proportional hazards–based extrapolation could be undertaken using the IPCW HR by fitting a parametric model to the observed experimental group's survival data (which is unaffected by treatment switching) and multiplying the hazard function by the inverse of the IPCW HR to obtain the control group's hazard function, from which the control group survivor function could be derived. This may produce a degree of error because an HR is applied to an independently fitted parametric model, but this error is likely to be minimal.

## RESULTS OF A SIMULATION STUDY

In a previous work, we conducted an extensive simulation study that evaluated the performance of treatment switching adjustment methods across a wide range of scenarios.[28] We used a joint longitudinal and survival model to simultaneously generate a time-dependent prognostic covariate and survival times. Parameter values were selected such that simulated survival times were reflective of the type of data often observed in metastatic cancer trials. We tested different levels of switching proportion, treatment effect, and censoring and different switching mechanisms. In each simulation, the true survival differences between treatment options were known, allowing us to apply each switching adjustment method and assess their performance with respect to bias, mean squared error, and coverage. Our results confirmed those found in another simulation study[7]—that is, the RPSFTM and IPE methods perform very well when the ''common treatment effect'' assumption holds, while simple methods produce very high levels of bias—but also provided evidence on the comparative performance of relevant methods in scenarios in which their key assumptions did not hold.[28]

We demonstrated that the IPCW method represented a substantial improvement compared to simple methods but produced higher bias than the RPSFTM and IPE methods when the ''common treatment effect'' assumption held.[28] This was likely to be due to the error associated with applying an observation-based method to a relatively small RCT dataset (with a sample size of 500) and was in line with findings previously reported by Howe and others.[29] Bias associated with the IPCW method became extremely high in scenarios in which the proportion of control group patients who switched treatments increased to approximately 90%, leaving approximately 20 patients in the control group who did not switch.[28] We also found that excluding a covariate that influenced the probability of treatment switching (thus violating the ''no unmeasured confounders'' assumption) only had a minimal impact on the bias produced by the method; however, this was likely to be due to the high level of correlation between the simulated prognostic covariates. The IPCW method resulted in substantially lower bias than the simple censoring method, which demonstrated the importance of the ''no unmeasured confounders'' assumption, as the IPCW reduces to simple censoring when all confounders are unmeasured.

In scenarios in which the treatment effect received by switchers was approximately 15% lower than the average effect received by patients initially randomized to the experimental group, we found that the RPSFTM, IPE, and IPCW methods produced similar levels of bias in their estimates of the treatment effect.[28] All produced important levels of bias, equivalent to approximately 5% to 10% of the treatment effect. In scenarios in which the treatment effect received by switchers was approximately 25% lower than the average effect received by patients initially randomized to the experimental group, the IPCW

method produced lower bias than the RPSFTM and IPE methods (which often produced bias of >10%), and in these scenarios, the ITT analysis often produced the least bias (0%–5%) if the treatment effect was relatively low (equivalent to an HR of ~0.75 in experimental group patients).[28] This is logical because in these scenarios, patients who switch receive very little benefit from the experimental treatment.

In addition to the ''standard'' treatment switching adjustment methods described so far, in our simulation study, we tested two ''2-stage'' methods: an SNM with g-estimation, and a simple 2-stage Weibull approach. The SNM performed poorly in our simulations, particularly when switching proportions were very high.[28] The simple Weibull model performed much better, producing relatively low bias across all scenarios. It generally produced lower bias and was much less sensitive to the switching proportion than the IPCW method, perhaps reflecting its lower data and modeling requirements. While the RPSFTM and IPE methods produced less bias than the 2-stage Weibull method when the ''common treatment effect'' assumption held, the opposite was true when that assumption was violated. The results associated with the 2-stage Weibull method should be interpreted with some caution because it was well suited to the switching mechanism incorporated within the simulation study; in particular, switching could only occur soon after disease progression. However, it is noteworthy that the switching mechanism that was simulated was similar to that observed in metastatic cancer trials, and thus, the good results associated with the 2-stage Weibull method should not be ignored. This method is worthy of consideration in situations in which treatment switching can only occur after an identifiable secondary baseline; switching occurs soon after that secondary baseline; data on important prognostic factors are available at that secondary baseline; and the RPSFTM, IPE, and IPCW methods seem inappropriate.

## PRACTICAL RECOMMENDATIONS

Based on knowledge of the theoretical assumptions and limitations associated with the treatment switching adjustment methods, the practicalities of their application in an economic evaluation context, and their performance in simulation studies, it is possible to make practical recommendations on how they should be used in future economic evaluations. Given the limitations associated with the switching adjustment methods, these recommendations cannot be entirely conclusive or specific, but given the current lack of understanding of these methods in the HTA arena, they remain useful to make. We would expect these recommendations to evolve with further research. The recommendations are presented in the form of an analysis framework in Figure 2.

Step 1 involves assessing the treatment switching mechanism. This should demonstrate whether and which adjustment methods are potentially applicable. For instance, it may become apparent whether data on relevant switching indicators were collected. The time at which patients became able to switch treatments is also important to determine. For step 2, the proportion of treatment switching should be assessed. If more than 90% of control group patients switch, the IPCW method is highly prone to bias, given a sample size in the region of 500. This is likely to be the case for most cancer clinical trials because sample sizes are rarely larger than the 500 (250 in each arm) tested in our simulation study. It is likely that the sample size would need to be substantially greater than 500 for the IPCW to produce unbiased results when the proportion of patients who switch is as high as 90%. Randomization-based methods are relatively less affected by high levels of switching and therefore should be given precedence (unless there is evidence of a strong time-dependent treatment effect or the comparator included in the RCT is active, rendering the standard counterfactual survival model inappropriate).

Step 3 involves drawing upon steps 1 and 2 and assessing the pivotal assumptions of each of the adjustment methods to further determine which may be potentially appropriate. For the RPSFTM and IPE algorithm, the ''common treatment effect'' assumption should be assessed. Survival models with the randomized group included as a covariate and a switching indicator variable may be used, but the potential bias associated with these should be recognized. Depending on the extent to which treatment switching occurred, log-cumulative hazard and quantile-quantile plots may remain useful for assessing the proportionality of hazards and the constancy of the acceleration factor over time. If patients with different stages of disease were randomized to the trial, the treatment effect in these subgroups should be investigated to offer further evidence on the ''common treatment effect'' assumption, although this may also be prone to bias due to switching. Given the limitations associated with assessing the ''common treatment effect'' assumption using trial data, external data sources should be sought, and expert opinion on the clinical and biological plausibility of

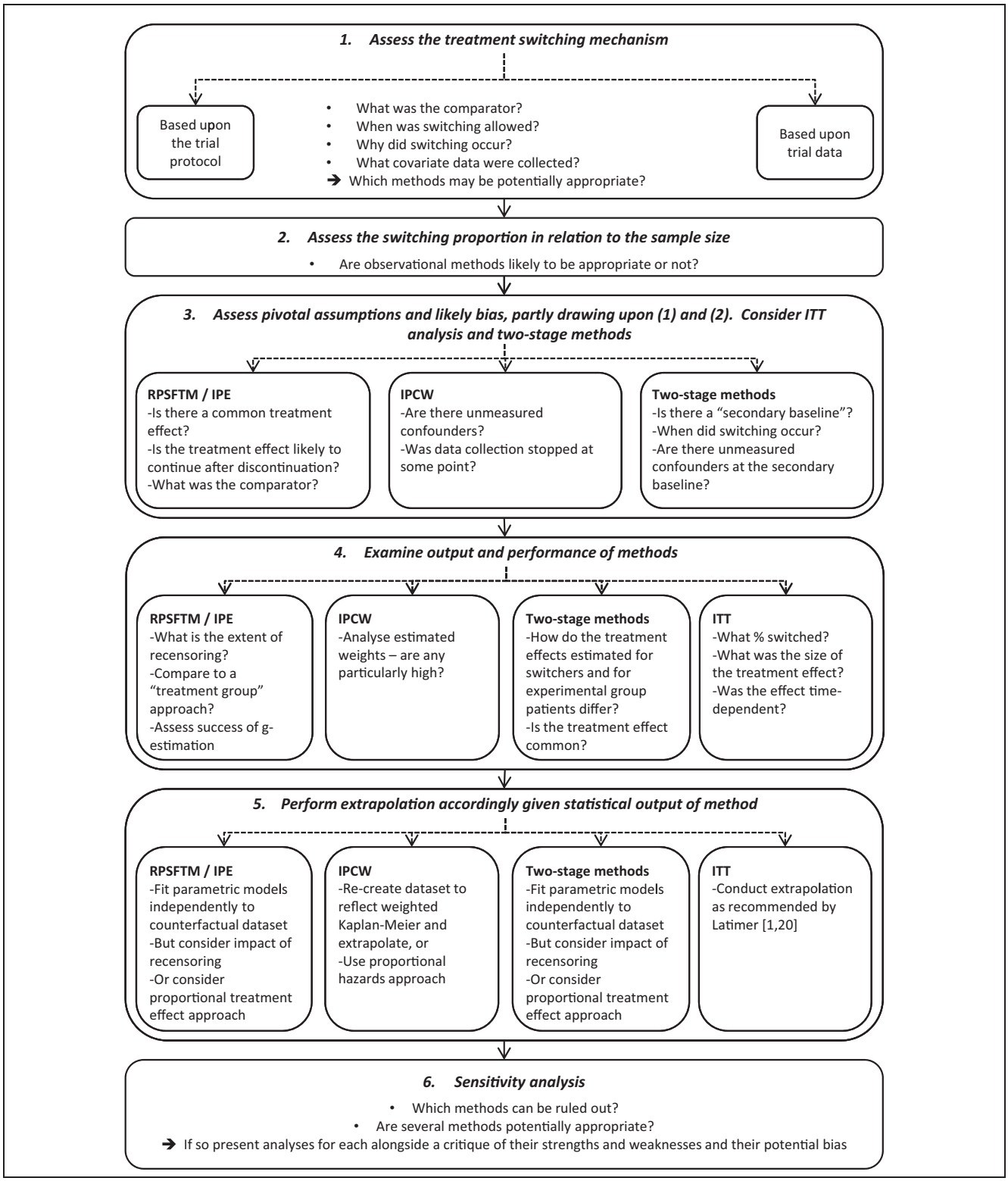


*Figure 2 Treatment switching analysis framework.*

the assumption must be considered. If these analyses suggest that the ''common treatment effect'' assumption holds, an RPSFTM or IPE approach may be used. However, consideration should also be given to the comparator included in the RCT (i.e., whether it is active or not) and the duration of the treatment effect (i.e., whether it is likely to be maintained to any extent after treatment discontinuation). If it is likely that the treatment effect may be maintained beyond treatment discontinuation, a ''treatment group'' application of the RPSFTM or IPE algorithm might be considered.

For the IPCW, the ''no unmeasured confounders'' assumption should be considered. The likelihood that data on important covariates were not collected should be informed by clinical expert opinion as well as an assessment of covariate data reported from other trials in similar disease areas. This alone is not sufficient to guarantee that the ''no unmeasured confounders'' assumption is satisfied because unknown confounders may exist. It is necessary to record all prognostic information that may have influenced decisions to switch; this includes the clinician's opinion on whether a patient is suitable for switching and patient circumstances and their preference for switching. Information on these factors is not routinely collected in RCTs. Combined with this, consideration should be given to whether the collection of covariate data stopped at any point during the trial (e.g., at the point of disease progression), as this restricts the applicability of the IPCW method. These issues should be considered in combination with those specified in steps 1 and 2.

When considering the use of 2-stage methods, the existence of an appropriate secondary baseline (such as disease progression) is pivotal. These will only exist if there is a time point before which treatment switching could not occur. If such a time point exists, 2-stage methods are possible to apply, but their potential bias will be related to how soon after this point switching occurs; if there are long delays until switching, the potential for bias associated with time-dependent confounding becomes important.

After applying the switching adjustment methods, step 4 involves a review of the output of the methods to help identify whether the methods are likely to have performed well. For the RPSFTM and IPE methods, this includes a consideration of the degree of recensoring and possibly a comparison of standard RPSFTM and IPE findings to results when these methods are applied on a ''treatment group'' basis in order to identify whether the treatment effect may have continued beyond treatment discontinuation. It is also important to assess the g-estimation output in order to identify the success with which the RPSFTM method has identified a unique treatment effect and whether the RPSFTM and IPE methods produce treatment effects that result in equal counterfactual survival times between randomized groups. For the IPCW, it is particularly important to assess the weights calculated for each patient over time; instances where certain patients are allocated particularly high weights are likely to lead to erroneous IPCW results. Outputs from 2-stage methods may be used to help determine the appropriateness of other methods; for instance, if the 2-stage methods produce estimates of the treatment effect in the switching patients that are (not) similar to the effect estimated for patients randomized to the experimental group, the RPSFTM/IPE methods may (not) be appropriate.

In tandem with a consideration of complex switching adjustment methods, a standard ITT analysis should be considered, as if other methods are likely to have performed poorly, the ITT analysis may provide the least bias. If the treatment effect is small (with an HR of ~0.75–1.00 in the experimental group) and there is evidence of switchers receiving a treatment effect that is around 15% lower than that received by experimental group patients, an ITT analysis is likely to be preferable to the IPCW and RPSFTM/IPE methods (although this will still contain bias). If the decrement in the treatment effect received by switchers is stronger, around 25%, the ITT analysis is even more likely to be preferable to the IPCW and RPSFTM/IPE methods unless the treatment effect is high (equivalent to an HR of ~0.50). Given the limitations associated with switching adjustment methods, the ITT analysis should always be presented. All other things being equal, in situations where switching proportions are low and/or the treatment effect is low and/or the treatment effect is likely to be much reduced in switchers, the ITT analysis may provide the least bias.

Step 5 addresses combining the adjustment methods with an extrapolation approach (if required) based on the statistical output of the applied adjustment method. Finally, when a preliminary analysis of trial data suggests that the choice of a preferable adjustment method is unclear, a sensitivity analysis should be undertaken to demonstrate the uncertainty associated with the methodology used.

## DISCUSSION

Treatment switching adjustment methods have often been used poorly and have been inadequately

described in economic evaluations. Our review of NICE TAs (summarized in Table 1 and discussed in Appendix A) demonstrates that while some potentially appropriate methods have been used, more often, simple methods that are highly prone to bias have been relied on. Where more complex, potentially appropriate methods such as the RPSFTM and IPCW have been used, discussion of these methods within the appraisal documents has been lacking, failing to consider their key limitations.[10,11] This is important because the application of switching adjustment methods within an economic model often drastically alters the estimated ICER. Through a consideration of the theoretical properties of available adjustment methods and the results of a simulation study, we have developed an analysis framework that can be used in future HTAs affected by switching to reduce the use of inappropriate and inconsistent methods.

Because the RPSFTM and IPCW methods work in very different ways and make very different assumptions, one is unlikely to always be better than the other. Trial and switching characteristics must be considered on a case-by-case basis to assess which switching adjustment method is likely to be valid. The IPCW has observational data origins, and its reliance on the ''no unmeasured confounders'' assumption represents a very important limitation that may be difficult to justify in an RCT setting. The RPSFTM and IPE methods are limited by the ''common treatment effect'' assumption, which may appear clinically implausible in situations where treatment switching occurs after disease progression. Previously unused, simple 2-stage methods should be considered, particularly in circumstances in which the RPSFTM, IPE, and IPCW methods are highly prone to bias. These require a suitable secondary baseline to be present but do not make the ''common treatment effect'' assumption and only require the ''no unmeasured confounders'' assumption to hold at the secondary baseline time point. However, simple 2-stage methods remain prone to time-dependent confounding, although this may be limited where switching occurs soon after the secondary baseline. In our simulation study, we chose a Weibull model when we tested this 2-stage approach, but in reality, a preferred model could be identified by examining the goodness of fit of a variety of accelerated failure time models to the control group postprogression dataset.

While our analysis framework attempts to enhance the probability that inappropriate adjustment methods are avoided, in some scenarios, no ''good'' methods are available. In situations where the ''common treatment effect'' assumption appears unreasonable and the proportion of patients who switch is very high (e.g., ~90% in a control group sample size in the region of 250 patients), the RPSFTM and IPE methods may not be appropriate, and the IPCW method is prone to high levels of bias. Very high switching proportions combined with small sample sizes are likely to cause 2-stage methods also to become prone to error and bias; although this was not demonstrated in our simulation study,[28] these methods should be used with caution in such circumstances. This reflects the current lack of suitable methods to address realistic scenarios, and hence, research into novel methods would be highly valuable. In addition, while our simulation study provided important evidence on the use of switching adjustment methods in realistic scenarios, running further scenarios with different treatment effects, switching proportions, and data-generating mechanisms would be useful.

Our analyses also demonstrate that the use of several treatment switching adjustment methods require the collection of suitable data in clinical trials. Data on patient characteristics that are prognostic and that are predictive of treatment switching are required at baseline and over time. If switching is to be permitted, clinical trialists should develop protocols that ensure that the required data are collected during the trial to enhance the likelihood that appropriate adjustments can be made for subsequent HTA analyses.

It is worth reiterating that the ITT analysis remains important even in the presence of treatment switching. If the novel treatment is found to be cost-effective under an ITT analysis, despite treatment switching, this may increase decision makers' confidence that it represents a cost-effective use of resources. In addition, when switchers are expected to receive a much lower treatment effect than patients randomized to the experimental treatment, an ITT analysis may result in relatively low bias.

We have focused on adjusting survival time estimates in the presence of treatment switching from the control treatment to the experimental treatment. In some circumstances, it may be desirable to also adjust for switching from the experimental treatment to the control treatment or for switching to other alternative therapies, although often, such switches may represent realistic treatment pathways that do not require adjustment within an economic evaluation context. The RPSFTM and IPE methods are designed to cope with treatment switching in either direction

(provided that the control treatment is placebo, or nonactive) but are not suitable when switching is to a third treatment. In such circumstances, a multiparameter RPSFTM would be required, but this has been shown to perform poorly in practice.[13,24,25] Theoretically, the IPCW and 2-stage methods could be adapted to adjust for switching in any direction to any treatment, with models being applied to different groups as appropriate. However, increasing the number of adjustments made to the observed dataset may further compound the data requirements associated with these methods, potentially rendering them prone to increasing bias.

It is important to note that other parameters included in an economic evaluation are likely to be affected by treatment switching. Where quality of life and cost data are collected within a clinical trial affected by switching, ITT analyses of these outcomes will be confounded. Aside from simply excluding the direct costs of treatments that were switched to,[10,11] we are unaware of attempts to adjust for the effects of switching on these outcomes in HTAs. The problem may not be as serious as for survival estimates; quality of life scores are often based on health states rather than the treatment group, and direct and indirect costs are often based on assumption or external sources[10,11]; yet, where trial data are used, the confounding represents an important issue. Switching to a beneficial treatment is likely to have an effect on quality of life and resource use, and failure to adjust for this may result in the economic evaluation providing biased estimates of the relative cost-effectiveness of the treatment. Methods to undertake such adjustments are available; when the mean outcome is of interest, a structural mean model may be suitable, and with repeated outcomes, a structural nested mean model may be appropriate[30–33]; however, further research on the use of these methods within an economic evaluation context would be extremely valuable. Our simulation study[28] only considered re-estimation of survival times, and so, in this sense, its scope was not sufficient to fully address the treatment switching issue in the context of an economic evaluation.

Finally, it is important to recognize that we have focused on the use of within-trial statistical methods to address the treatment switching problem rather than methods that make use of external data. Often, suitable external data (e.g., external trials not confounded by switching, or registry data) will not be available, but where it is, methods to formally synthesize data would have value. This is particularly important because the statistical adjustment methods focused on in this article often produce highly uncertain estimates of the treatment effect, with wide confidence intervals, reflecting the uncertainty associated with estimating counterfactual survival times and treatment effects. Related to this, we have considered only situations where patient-level data are available; research into the potential for making adjustments for switching without such data, particularly for use within indirect comparisons, would be of high value.

## CONCLUSIONS

It is clear that treatment switching is an important factor in a substantial proportion of HTAs, particularly in the oncology setting. Our article offers recommendations on the use of treatment switching adjustment methods that, if used, enhance the likelihood that appropriate methods are identified and used in future HTAs. In addition, we recommend that clinical trialists ensure that suitable data are collected within RCTs to allow switching adjustment methods to be applied.

## REFERENCES

1. Center for Drug Evaluation and Research, Center for Biologics Evaluation and Research, eds. Guidance for Industry: Clinical Trial Endpoints for the Approval of Cancer Drugs and Biologics. Silver Spring (MD): US Department of Health and Human Services, Food and Drug Administration; 2007.

2. Committee for Medicinal Products for Human Use. Appendix 1 to the Guideline on the Evaluation of Anticancer Medicinal Products in Man (CHMP/EWP/205/95 REV.3): Methodological Considerations for Using Progression-Free Survival (PFS) as Primary Endpoint in Confirmatory Trials for Registration. London: European Medicines Agency; 2006. Available from: http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/12/WC500017736.pdf.

3. National Institute for Health and Care Excellence (NICE). Guide to the Methods of Technology Appraisal. London: NICE; 2008. Available from: http://www.nice.org.uk/media/B52/A7/TAMethodsGuideUpdatedJune2008.pdf.

4. Briggs A, Claxton K, Sculpher M. Decision Modelling for Health Economic Evaluation. New York: Oxford University Press Inc.; 2006.

5. Gold MR, Siegel JE, Russell LB, Weinstein MC. Cost-Effectiveness in Health and Medicine. New York: Oxford University Press Inc.; 1996.

6. Canadian Agency for Drugs and Technologies in Health (CADTH). Guidelines for the Economic Evaluation of Health Technologies. 3rd ed. Ottawa: CADTH; 2006.

7. Morden JP, Lambert PC, Latimer NR, Abrams KR, Wailoo AJ. Assessing methods for dealing with treatment switching in randomised controlled trials: a simulation study. BMC Med Res Methodol. 2011;11:4.

8. Lee Y, Ellenberg J, Hirtz D, Nelson K. Analysis of clinical trials by treatment actually received: is it really an option? Stat Med. 1991;10:1595–605.

9. Horwitz R, Horwitz S. Adherence to treatment and health outcomes. Arch Intern Med. 1993;153:1863–8.

10. National Institute for Health and Care Excellence (NICE). Pazopanib for the First-Line Treatment of Metastatic Renal Cell Carcinoma: TA 215. London: NICE; 2011.

11. National Institute for Health and Care Excellence (NICE). Everolimus for the Second-Line Treatment of Advanced Renal Cell Carcinoma: TA 219. London: NICE; 2011.

12. Robins JM, 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(3):779–88.

13. Robins JM, Greenland S. Adjusting for differential rates of prophylaxis therapy for PCP in high-dose versus low-dose AZT treatment arms in an AIDS randomized trial. J Am Statist Assoc. 1994; 89(427):737–49.

14. Yamaguchi T, Ohashi Y. Adjusting for differential proportions of second-line treatment in cancer clinical trials, part I: structural nested models and marginal structural models to test and estimate treatment arm effects. Stat Med. 2004;23(13): 1991–2003.

15. Hernan MA, Brumback B, Robins JM. Marginal structural models to estimate the joint causal effect of nonrandomized treatments. J Am Statist Assoc. 2001;96(454):440–8.

16. Robins JM. Marginal structural models versus structural nested models as tools for causal inference. In: Halloran ME, Berry D, eds. Statistical Models in Epidemiology: The Environment and Clinical Trials. New York: Springer-Verlag; 1999. p 95–134.

17. Bond M, Hoyle M, Moxham T, Napier M, Anderson R. The Clinical and Cost-Effectiveness of Sunitinib for the Treatment of Gastrointestinal Stromal Tumours: A Critique of the Submission from Pfizer. Peninsula Technology Assessment Group, Universities of Exeter and Plymouth, commissioned by the NHS R&D HTA Programme on behalf of the National Institute for Clinical Excellence. 2009. Available from: http://www.nice.org.uk/nicemedia/live/12040/43430/43430.pdf.

18. Robins JM, Tsiatis AA. Correcting for noncompliance in randomized trials using rank preserving structural failure time models. Commun Stat Theory Methods. 1991;20(8):2609–31.

19. Hernan MA, Robins JM. Instruments for causal inference: an epidemiologist's dream? Epidemiology. 2006;17:360–72.

20. Senn SJ. Covariate imbalance and random allocation in clinical trials. Stat Med. 1989;8(4):467–75.

21. Hampson LV, Metcalfe C. Incorporating prognostic factors into causal estimators: a comparison of methods for randomised controlled trials with a time-to-event outcome. Stat Med. 2012; 31(26):3073–88.

22. Branson M, Whitehead J. Estimating a treatment effect in survival studies in which patients switch treatment. Stat Med. 2002; 21(17):2449–63.

23. Latimer NR. Survival analysis for economic evaluations alongside clinical trials—extrapolation with patient-level data: inconsistencies, limitations and a practical guide. Med Decis Making. 2013;33(6):743–54.

24. White IR, Babiker AG, Walker S, Darbyshire JH. Randomization-based methods for correcting for treatment changes: examples from the Concorde trial. Stat Med. 1999;18(19):2617–34.

25. Yamaguchi T, Ohashi Y. Adjusting for differential proportions of second-line treatment in cancer clinical trials, part II: an application in a clinical trial of unresectable non-small-cell lung cancer. Stat Med. 2004;23(13):2005–22.

26. Guyot P, Welton NJ, Ouwens MJ, Ades AE. Survival time outcomes in randomized, controlled trials and meta-analyses: the parallel universes of efficacy and cost-effectiveness. Value Health. 2011;14:640–6.

27. Guyot P, Ades AE, Ouwens MJNM, Welton NJ. Enhanced secondary analysis of survival data: reconstructing the data from published Kaplan-Meier survival curves. BMC Med Res Methodol. 2012;12:9.

28. Latimer N, Abrams KR, Lambert PC, et al. Adjusting for treatment switching in randomised controlled trials: a simulation study. University of Sheffield Health Economics and Decision Science Discussion Paper. 2013. No. 13/06. Available from: http://www.shef.ac.uk/polopoly_fs/1.259501!/file/HEDSDP1306.pdf.

29. Howe CJ, Cole SR, Chmiel JS, Munoz A. Limitation of inverse probability-of-censoring weights in estimating survival in the presence of strong selection bias. Am J Epidemiol. 2011;173(5):569–77.

30. Goetghebeur E, Lapp K. The effect of treatment compliance in a placebo- controlled trial: regression with unpaired data. Appl Statist. 1997;46:351–64.

31. Fischer-Lapp K, Goetghebeur E. Practical properties of some structural mean analyses of the effect of compliance in randomized trials. Control Clin Trials. 1999;20:531–46.

32. Robins JM. Correcting for non-compliance in randomized trials using structural nested mean models. Commun Stat Theory Methods. 1994;23:2379–412.

33. White IR. Uses and limitations of randomization-based efficacy estimators. Stat Methods Med Res. 2005;14(4):327–47.

34. Vermorken JB, Mesia R, Rivera F, et al. Platinum-based chemotherapy plus cetuximab in head and neck cancer. N Engl J Med. 2008;359:1116–27.

35. Roche Products Ltd. Achieving Clinical Excellence in the Treatment of Relapsed Non-Small Cell Lung Cancer: Tarceva (Erlotinib) NICE STA Submission. Welwyn Garden City (UK): Roche Products Ltd.; 2006.

36. Lewis R, Bagnall AM, Forbes C, et al. A Rapid and Systematic Review of the Clinical Effectiveness and Cost Effectiveness of Trastuzumab for Breast Cancer. NHS Centre for Reviews and Dissemination Report commissioned by the NHS R&D HTA Programme on behalf of the National Institute for Clinical Excellence. 2001. Available from: http://www.nice.org.uk/nicemedia/live/11445/32317/32317.pdf.

37. National Institute for Health and Care Excellence (NICE). Final Appraisal Determination: Imatinib for the Treatment of Unresectable and/or Metastatic Gastro-intestinal Stromal Tumours: TA 86. London: NICE; 2004.

38. Janssen-Cilag Ltd. STA Submission to NICE: Velcade (Bortezomib) for the Treatment of Multiple Myeloma Patients at First Relapse. High Wycombe (UK): Janssen-Cilag Ltd.; 2006.

39. Hoyle M, Rogers G, Garside R, Moxham T, Stein K. 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. Peninsula Technology Assessment Group, Universities of Exeter and Plymouth, commissioned by the NHS R&D HTA Programme on behalf of the National Institute for Clinical Excellence. 2008. Available from: http://www.hta.ac.uk/erg/reports/1746.pdf.

40. Hind D, Tappenden P, Tumur I, Eggington S, Sutcliffe P, Ryan A. The Use of Irinotecan, Oxaliplatin and Raltitrexed for the Treatment of Advanced Colorectal Cancer: Systematic Review and Economic Evaluation (Review of Guidance No. 33). Addendum: Economic Evaluation of Irinotecan and Oxaliplatin for the Treatment of Advanced Colorectal Cancer. School of Health and Related Research, University of Sheffield, commissioned by the NHS R&D HTA Programme on behalf of the National Institute for Clinical Excellence. 2005. Available from: http://www.nice.org.uk/nicemedia/live/11561/33130/33130.pdf.

41. Merck Serono Ltd. Single Technology Appraisal Submission: Erbitux (Cetuximab) for the First-Line Treatment of Metastatic Colorectal Cancer. London: Merck Serono Ltd.; 2008.

42. National Institute for Health and Care Excellence (NICE). Guidance on the Use of Trastuzumab for the Treatment of Advanced Breast Cancer: NICE Technology Appraisal Guidance No. 34. London: NICE; 2002.

43. Robins JM. Structural nested failure time models. In: Andersen PK, Keiding N, section eds. Survival Analysis, Armitage P, Colton T, eds. The Encyclopedia of Biostatistics. Chichester (UK): John Wiley and Sons; 1998. p 4372–89.

44. Mark SD, Robins JM. A method for the analysis of randomized trials with compliance information: an application to the multiple risk factor intervention trial. Control Clin Trials. 1993;14(2): 79–97.