

**NICE DSU TECHNICAL SUPPORT DOCUMENT 16:
ADJUSTING SURVIVAL TIME ESTIMATES IN THE
PRESENCE OF TREATMENT SWITCHING**

REPORT BY THE DECISION SUPPORT UNIT

July 2014

Nicholas R Latimer¹

Keith R Abrams²

¹ School of Health and Related Research, University of Sheffield, UK

² Department of Health Sciences, University of Leicester, UK

Decision Support Unit, ScHARR, University of Sheffield, Regent Court, 30 Regent Street
Sheffield, S1 4DA

Tel (+44) (0)114 222 0734

E-mail dsuadmin@sheffield.ac.uk

Twitter: [@NICE_DSU](https://twitter.com/NICE_DSU)

ABOUT THE DECISION SUPPORT UNIT

The Decision Support Unit (DSU) is a collaboration between the Universities of Sheffield, York and Leicester. We also have members at the University of Bristol, London School of Hygiene and Tropical Medicine and Brunel University.

The DSU is commissioned by The National Institute for Health and Care Excellence (NICE) to provide a research and training resource to support the Institute's Technology Appraisal Programme. Please see our website for further information www.nicedsu.org.uk

ABOUT THE TECHNICAL SUPPORT DOCUMENT SERIES

The NICE Guide to the Methods of Technology Appraisalⁱ is a regularly updated document that provides an overview of the key principles and methods of health technology assessment and appraisal for use in NICE appraisals. The Methods Guide does not provide detailed advice on how to implement and apply the methods it describes. This DSU series of Technical Support Documents (TSDs) is intended to complement the Methods Guide by providing detailed information on how to implement specific methods.

The TSDs provide a review of the current state of the art in each topic area, and make clear recommendations on the implementation of methods and reporting standards where it is appropriate to do so. They aim to provide assistance to all those involved in submitting or critiquing evidence as part of NICE Technology Appraisals, whether manufacturers, assessment groups or any other stakeholder type.

We recognise that there are areas of uncertainty, controversy and rapid development. It is our intention that such areas are indicated in the TSDs. All TSDs are extensively peer reviewed prior to publication (the names of peer reviewers appear in the acknowledgements for each document). Nevertheless, the responsibility for each TSD lies with the authors and we welcome any constructive feedback on the content or suggestions for further guides.

Please be aware that whilst the DSU is funded by NICE, these documents do not constitute formal NICE guidance or policy.

Dr Allan Wailoo

Director of DSU and TSD series editor.

ⁱ National Institute for Health and Care Excellence. Guide to the methods of technology appraisal, 2013 (updated 2013), London.

Acknowledgements

The authors thank Paul Lambert, Michael Crowther, James Morden, Allan Wailoo, Ron Akehurst and Mike Campbell for valuable contributions made to work that has contributed to this document.

The DSU thanks Ian White, Andrew Briggs, Warren Cowell, Paul Tappenden and Melinda Goodall for reviewing this document.

The production of this document was funded by the National Institute for Health and Care Excellence (NICE) through its Decision Support Unit. The views, and any errors or omissions, expressed in this document are of the authors only. NICE may take account of part or all of this document if it considers it appropriate, but it is not bound to do so.

KRA is partly supported by the UK National Institute for Health Research (NIHR) as a Senior Investigator (NI-SI-0508-10061).

This report should be referenced as follows:

Latimer NR, Abrams KR. NICE DSU Technical Support Document 16: Adjusting survival time estimates in the presence of treatment switching. (2014)

Available from <http://www.nicedsu.org.uk>

Competing interests

NRL has undertaken consultancy for Amgen, Astellas, AstraZeneca, Bayer, GSK, Novartis, Pfizer, and Sanofi Aventis.

KRA has received honoraria from Allergan, AstraZeneca, GSK, Janssen, Novartis, Novo Nordisk and Roche, and has acted as a paid consultant to Amaris, Creativ-Ceutical, OptumInsight and PRMA.

EXECUTIVE SUMMARY

Treatment switching can occur when patients in the control group of a clinical trial are allowed to switch onto the experimental treatment at some point during follow-up. Switching is common in clinical trials of cancer treatments and can also occur in trials of treatments for other diseases. Generally switching is permitted when the new intervention has been shown to be effective in interim analyses (often based upon an outcome measure such as time to disease progression), and it is deemed unethical to deny treatment to control group patients. Licensing bodies such as the United States Food and Drug Administration (FDA) and the European Medicines Agency (EMA), may accept progression free survival (PFS) as a primary endpoint for drug approval – reducing the incentives to maintain trial randomisation beyond disease progression.

When switching occurs, an “intention to treat” (ITT) analysis – whereby the data are analysed according to the arms to which patients were randomised – of the overall survival (OS) advantage associated with the new treatment will be biased: If control group patients switch treatments and benefit from the new treatment the OS advantage of the new treatment will be underestimated. For interventions that impact upon survival, health technology assessment (HTA) bodies such as the National Institute for Health and Care Excellence (NICE) require that economic evaluations consider a lifetime horizon. This is problematic in the presence of treatment switching, because standard ITT analyses are likely to be inappropriate.

Various statistical methods are available to adjust survival estimates in the presence of treatment switching, but each makes important assumptions and is subject to limitations. “Simple” adjustment methods such as censoring switchers at the point of switch, or excluding them entirely from the analysis, are highly prone to selection bias because switching is likely to be associated with prognosis. More complex adjustment methods, which are theoretically unbiased given certain assumptions are satisfied, are also available. Rank Preserving Structural Failure Time Models (RPSFTM) and the Iterative Parameter Estimation (IPE) algorithm represent randomisation-based methods for estimating counterfactual survival times (i.e. survival times that would have been observed in the absence of switching). The Inverse Probability of Censoring Weights (IPCW) method represents an observational-based approach, whereby data for switchers are censored at the point of switch and remaining observations are weighted with the aim of removing any censoring-related selection bias.

These methods all make important limiting assumptions – for instance the RPSFTM and the IPE algorithm rely critically on the “common treatment effect” assumption – that is, the treatment effect received by switchers must be the same (relative to the time the treatment is taken for) as the treatment effect received by patients initially randomised to the experimental group. This may not represent a valid assumption when patients who switch only receive the experimental treatment when their disease has progressed. Observational-based adjustment methods (such as the IPCW) are reliant on the “no unmeasured confounders” assumption – that is, data must be available on baseline and time-dependent variables that predict both treatment switching and prognosis.

This Technical Support Document (TSD) introduces the RPSFTM, IPE, IPCW and other adjustment methods that may be used in the presence of treatment switching. The key assumptions and limitations associated with each method are described, and the use of these in past NICE technology appraisals and their performance in simulation studies is reviewed. Based upon this, advice is offered in the form of an analysis framework, to help analysts determine adjustment methods that are likely to be appropriate on a case-by-case basis. Importantly, no single method will be optimal in all circumstances – the performance of alternative methods is dependent upon the characteristics of the trial to which they are applied. For instance, the IPCW method is highly prone to error when a very large proportion of control patients (greater than approximately 90%, in a trial with sample size 500) switch onto the experimental treatment. RPSFTM and IPE methods are sensitive to the “common treatment effect” assumption, but the importance of this sensitivity depends upon the size of the treatment effect observed in the trial in question. Novel two-stage methods appear to represent a valid alternative adjustment approach, but are only applicable when switching can only occur after a specific disease-related time-point (such as disease progression). Given the limitations associated with the adjustment methods, the ITT analysis should always be presented. Analysts should consider in detail the characteristics of the trial, the switching mechanism, the treatment effect, data availability and adjustment method outputs when determining and justifying appropriate adjustment methods. In addition to this, at the trial planning stage, researchers should take account of the data requirements of switching adjustment methods, if switching is to be permitted during the trial, or is thought likely to occur.

CONTENTS

1. INTRODUCTION	8
2. TREATMENT SWITCHING – THE PROBLEM.....	11
3. TREATMENT SWITCHING ADJUSTMENT METHODS	14
3.1 SIMPLE METHODS	14
3.1.1 <i>Intention-to-treat analysis</i>	14
3.1.2 <i>Per protocol analysis – excluding and censoring switchers</i>	14
3.1.3 <i>Including costs of the treatment switched to</i>	14
3.1.4 <i>Modelling based only on PFS</i>	15
3.1.5 <i>Applying the same risk of death upon disease progression</i>	16
3.1.6 <i>Assumed equal OS for the two treatment groups</i>	17
3.1.7 <i>Using sequencing models</i>	17
3.2 COMPLEX METHODS	18
3.2.1 <i>Inverse Probability of Censoring Weights</i>	18
3.2.2 <i>Rank Preserving Structural Failure Time Model</i>	19
3.2.3 <i>Iterative Parameter Estimation algorithm</i>	20
3.2.4 <i>Alternative “two-stage” methods</i>	21
3.2.5 <i>Using external data</i>	22
3.3 APPLICATION TO ECONOMIC EVALUATIONS	23
3.3.1 <i>Theoretical limitations</i>	23
3.3.2 <i>Practical limitations</i>	25
4. SIMULATION STUDIES	26
5. REVIEW OF SWITCHING ADJUSTMENT METHODS USED IN.....	32
NICE TAs	32
5.1 EXTERNAL DATA	35
5.2 REVIEW CONCLUSIONS	37
6. METHODOLOGICAL AND PROCESS GUIDANCE	38
7. DISCUSSION.....	44
8. CONCLUSIONS.....	47
9. REFERENCES	48
APPENDIX A: COMPLEX SWITCHING ADJUSTMENT METHODS	53

TABLES & FIGURES

Table 1: Methods used to account for switching in NICE technology appraisals (2000 – 2009).....	34
---	----

Figure 1: The potential impact of treatment switching illustrated	13
---	----

Figure 2: Treatment switching analysis framework.....	40
---	----

Abbreviations and definitions

AF	Acceleration Factor
AG	Assessment Group
DSU	Decision Support Unit
EMA	European Medicines Agency
ERG	Evidence Review Group
FAD	Final Appraisal Determination
FDA	United States Food and Drug Administration
GIST	Gastro-intestinal Stromal Tumours
HR	Hazard Ratio
HTA	Health Technology Assessment
ICER	Incremental Cost Effectiveness Ratio
IPCW	Inverse Probability of Censoring Weights
IPE	Iterative Parameter Estimation
ITT	Intention To Treat
IV	Instrumental Variables
MSM	Marginal Structural Model
MTA	Multiple Technology Appraisal
NICE	National Institute for Health and Clinical Excellence
OS	Overall Survival
PFS	Progression Free Survival
PPS	Post Progression Survival
QALY	Quality Adjusted Life Year
RCC	Renal Cell Carcinoma
RCT	Randomised Controlled Trial
RPSFTM	Rank Preserving Structural Failure Time Model
SNM	Structural Nested Model
STA	Single Technology Appraisal
TA	Technology Appraisal
TSD	Technical Support Document
WKM	Weighted Kaplan-Meier

1. INTRODUCTION

Interventions that impact upon survival form a high proportion of the treatments appraised by the National Institute for Health and Care Excellence (NICE). A previous Technical Support Document (TSD) has provided recommendations for the extrapolation of survival data using patient-level data.¹ However, separate from the problem of extrapolation, survival data collected in clinical trials, particularly in the setting of metastatic cancer, are often confounded by treatment switching. This prevents a standard intention-to-treat (ITT) analysis from addressing the decision problem faced within the economic evaluation.²

In an economic evaluation, the decision problem generally requires the comparison of a state of the world in which the new therapy is available and is provided to a cohort of indicated patients, to a state of the world in which it is not. In this TSD treatment switching is defined as the switch from control treatment to experimental treatment by patients randomised to the control group of a Randomised Controlled Trial (RCT). Some authors term this “treatment crossover” rather than “treatment switching”; here we have used the term “switching” because “crossover” may evoke crossover trials, which are a different entity. In the presence of treatment switching the control group is contaminated – it no longer represents the state of the world in which the new treatment is not available. To address the economic evaluation decision problem, adjustments must be made to the observed data in order to obtain a more robust estimate of the relative benefit of the intervention compared to the control.

Treatment switching is common in trials of oncology treatments, and can be an issue in other non-cancer areas. It has had an important impact in several NICE technology appraisals (TAs), as shown in Section 5 of this report. Switching is prevalent for both ethical and practical reasons. Ethically, when there are no other non-palliative 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 (in some cases) of progression-free survival (PFS) as a primary endpoint for drug regulatory approval by agencies such as the United States Food and Drug Administration (FDA) and the European Medicines Agency (EMA).^{3,4} RCTs of cancer treatments are therefore often powered to investigate differences in PFS rather than overall survival (OS) and there is less motivation for pharmaceutical

companies to ensure that randomised groups are maintained beyond disease progression (hence treatment switching may be allowed beyond this point). However, whilst showing an OS advantage may not be essential for obtaining marketing authorisation, a lifetime horizon is generally advocated in economic evaluations, especially for interventions that impact upon survival; this is recommended in the NICE Guide to the Methods of Technology Appraisal⁵ and in methodological guidance in other health care jurisdictions.⁶⁻⁸ For this reason the issue of treatment switching is often regarded as one related to economic evaluation and in this TSD it is the economic evaluation context that we address. However, treatment switching is important for the interpretation of the clinical evidence more broadly because the aim of adjustment analyses is to obtain more accurate estimates of the clinical benefit associated with the new treatment.

In this TSD, first we set out the treatment switching problem in the context of economic evaluation. We then summarise the key switching adjustment methods. Rather than providing mathematical formulae, we outline their pivotal assumptions and limitations as well as their practical applicability in an economic evaluation context. More detail on the methodological theory are given in Appendix A, and relevant technical papers are referred to. Results of simulation studies are then summarised in order to demonstrate the bias that might be expected to be associated with the different switching adjustment methods in a range of different scenarios. Next, we summarise how switching adjustment methods have been used in an HTA context based upon a review of NICE TAs. Finally, we offer guidance on the use of adjustment methods in the form of an analysis framework and provide discussion around this. Much of the material contained within this TSD is covered in a related paper recently published by *Medical Decision Making*.² Where additional research has been completed since the publication of that paper, it is captured within this TSD.

This TSD is limited to a discussion of methods that may be used to adjust survival time estimates in the presence of treatment switching – we do not consider the adjustment of other outcomes that might be affected by treatment switching. For instance, costs and quality of life scores collected within an RCT and attributed to randomised groups will be subject to confounding where switching is present. Aside from simply excluding the costs of treatments that were switched to, or only considering quality of life scores in non-switchers, we are unaware of attempts to adjust for the effects of switching on these outcomes in HTA, although structural mean models could potentially be used for this purpose.⁹⁻¹² The problem

may not be as serious as for survival estimates as quality of life scores are often based upon health states rather than treatment group, and direct and indirect costs are often based upon assumption or external data; however further research in these areas would be valuable.

In addition, this TSD focuses upon treatment switching from the control group onto the experimental treatment – we do not consider in detail situations in which experimental group patients switch onto the control treatment, or where patients randomised to either group receive other post-study 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 an appraisal of the relevant economic evaluation decision problem is still possible. If a patient randomised to the experimental group of a trial discontinues the novel therapy and subsequently receives a standard treatment (either that received in the control group or a separate standard treatment) this is likely to have occurred due to 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 – we would simply capture them within our analysis. Similarly, if patients randomised to the control (or experimental) group received post-study therapies that do not include the experimental treatment, this may reflect 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 post-study therapies this may reflect appropriate treatment pathways given the initial treatment. In each case, a judgement is required as to whether the treatment pathways observed represent realistic treatment patterns. Unless it is judged that this is not the case, it would be inappropriate to adjust for these differences in the economic analysis. The methods discussed in Section 3 of this TSD are not limited to the treatment switching definition that we use – they (or variations of them) can be used to address other forms of treatment changes.

It is important to note that this TSD does not attempt to provide prescriptive advice on exactly which methods should be used to adjust for treatment switching given different trial characteristics. Such guidance is not possible, because identifying the method that is likely to produce least bias is a function of several different factors and interactions of these (such as switching proportion, trial sample size, data collected, switching mechanism, magnitude of

treatment effect). It is not possible to cover every possible combination of these factors. Instead, we provide an analysis framework that will enhance the likelihood that suitable adjustment methods are identified on a case-by-case basis.

This TSD focuses on situations where patient-level data are available, because this is essential in order to use the switching adjustment methods detailed in Section 3. Hence, the TSD is particularly relevant to those preparing sponsor submissions to NICE. However, undertaking and reporting switching adjustment analysis as suggested in this TSD will also enable Assessment Groups (AGs) to critique sponsor submissions more effectively and in circumstances where patient-level data are provided to AGs, we recommend that they follow the processes outlined here. Research is ongoing regarding methods for adjusting for treatment switching in the absence of patient-level data.¹³

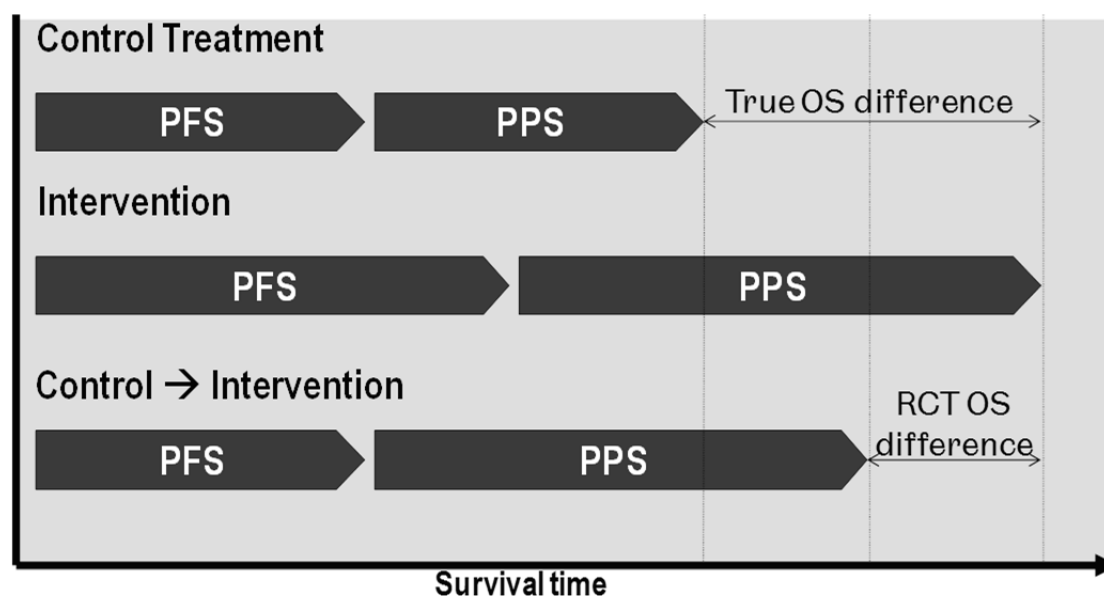
2. TREATMENT SWITCHING – THE PROBLEM

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.² Treatment switching causes a mismatch between what has been studied in the clinical trial and the economic analysts’ decision problem – an ITT analysis (a comparison of treatment groups as randomised) becomes insufficient to address the decision problem.

In this TSD 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 two rows (“Control Treatment” and “Intervention”) illustrate the “perfect” trial, where no treatment switching occurs. Survival time is on the x-axis, and in this example the new intervention extends PFS and post progression 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 ITT analysis (labelled “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 switching patients benefit to any extent from the new intervention, some bias will exist. An economic evaluation that relies upon this ITT analysis would produce inaccurate cost-effectiveness results (in this case the incremental cost effectiveness ratio (ICER) would be over-estimated) and inappropriate resource allocation decisions may be made.

Figure 1: The potential impact of treatment switching illustrated



Notes: PFS = Progression Free Survival; PPS = Post Progression Survival; OS = Overall Survival; RCT = Randomised Controlled Trial

The problems associated with treatment switching have been recognised in the NICE Guide to the Methods of Technology Appraisal,⁵ which states:

“In RCTs, participants randomised to the control group are sometimes allowed to switch treatment group and receive the active intervention. In these circumstances, when intention-to-treat analysis is considered inappropriate, statistical methods that adjust for treatment switching can also be presented. Simple adjustment methods such as censoring or excluding data from patients who crossover should be avoided because they are very susceptible to selection bias. The relative merits and limitations of the methods chosen to explore the impact of switching treatments should be explored and justified with respect to the method chosen and in relation to the specific characteristics of the data set in question. These characteristics include the mechanism of crossover used in the trial, the availability of data on baseline and time-dependent characteristics, and expectations around the treatment effect if the patients had remained on the treatment to which they were allocated.” [p.46, NICE Guide to the Methods of Technology Appraisal, 2013]

While the NICE Guide clearly recognises the limitations of ITT analyses and simple exclusion and censoring methods in the presence of treatment switching, it is not prescriptive with regards to the switching adjustment methods that should be used. This TSD provides further support with regard to identifying suitable adjustment methods.

3. TREATMENT SWITCHING ADJUSTMENT 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 HTA, as demonstrated by our review of NICE technology appraisals (TAs) presented in Section 5. First we discuss the key assumptions of the key methods, before considering their theoretical and practical limitations with respect to their incorporation within an economic model. We focus on the key principles of the methods rather than their mathematics – though further details on the more complex methods are provided in Appendix A.

3.1 SIMPLE METHODS

3.1.1 *Intention-to-treat analysis*

An ITT analysis does not attempt to adjust for treatment switching, but represents the standard analysis undertaken alongside RCTs. Groups are compared as randomised, and thus the randomisation-balance of the trial is respected. The ITT analysis represents a valid comparison of randomised 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 onto the novel therapy.

3.1.2 *Per protocol analysis – excluding and censoring switchers*

These approaches involve either excluding data from patients that switch, or censoring these at the point of the switch. Such analyses are prone to selection bias through informative censoring or exclusion because the randomisation balance between groups is broken if switching is associated with prognostic patient characteristics – for instance, if patients with poor prognosis are more likely to switch.^{14,15} 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 upon patient characteristics rather than being random.

3.1.3 *Including costs of the treatment switched to*

This approach represents an accurate economic evaluation of the RCT because it models exactly what happened in the trial; survival estimates are not adjusted for treatment

switching, but the cost of the treatments switched to are included in the analysis. This might be described as a “full ITT” analysis. The usefulness of the technique for use in HTA is uncertain, given the economic evaluation decision problem. As discussed in Section 1, the aim of an economic evaluation of a new intervention is typically to compare a state of the world where the new intervention exists to one in which it does not, and simply analysing the trial data as observed and allocating costs to patients who switch does not satisfy this aim – it trades internal consistency for external validity. An economic evaluation that incorporated the costs of the treatment switched to may dilute the bias associated with an ITT analysis that did not incorporate the switching costs, 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. It is likely that switchers are selected based upon prognosis, and they may also have a reduced capacity to benefit: these issues may cause the ICER to be importantly different in switchers compared to patients randomised 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.

3.1.4 Modelling based only on PFS

Where switching is only permitted after disease progression, data on PFS are not confounded. Hence, an option for the economic modeller may be to base the analysis on PFS only, thereby excluding data on post-progression survival. However, this does not mean that there is an implicit assumption that the new intervention only affects PFS. Rather, the method implies an assumption that the absolute QALY gain associated with the extension of PFS is exactly equivalent to the absolute QALY gain if OS had also been modelled, thus assuming that post-progression survival is identical in the two treatment groups and that PFS is an exact surrogate for OS. The initial conclusion may be that this method is likely to underestimate the cost-effectiveness of the new intervention: If the new intervention increases the duration of PFS, and if there is a link between PFS and OS, or if the new intervention has any independent effect on OS, then modelling based only upon PFS will underestimate cost-effectiveness. However, this is not necessarily the case. Modelling based only upon PFS essentially assumes that upon disease progression a patient dies – no more costs are incurred and no more QALYs are obtained. This is important because additional QALYs (and costs) are accrued after disease progression, and indeed if the absolute effect of the new treatment on OS is smaller than that on PFS then modelling based only on PFS may overestimate the cost-effectiveness of the new intervention. When considering this, it is important to note the

important distinction between absolute and relative effects. The relative effect of a new treatment may be lower for OS than PFS, but this does not necessarily mean that the absolute difference in OS will also be lower. Because OS is a longer time period than PFS an absolute difference in OS that is the same (or greater) than the absolute difference in PFS can be achieved with a worse (higher) hazard ratio. Therefore it is clear that economic analyses that only include PFS could lead to underestimation or overestimation of the cost-effectiveness of the new intervention. The NICE Guide to the Methods of Technology Appraisal states that a lifetime horizon should be modelled for treatments that are likely to impact upon survival,⁵ and thus modelling only PFS is likely to be inadequate.

3.1.5 Applying the same risk of death upon disease progression

This approach is an extension to the method of basing the economic analysis only on PFS is to model OS (see Section 3.1.4), but would assume that once a patient has experienced disease progression, their risk of death is the same whether they were randomised to the control group or the intervention group. Using this technique will mean that the absolute difference in OS for the two treatments will be similar to the absolute difference in PFS.

Essentially, this method assumes that the treatment effect of the new intervention is of limited duration – it lasts only until disease progression. After this point, there is no additional gain to having been treated with the new intervention. On the other hand, the risk of death is not greater in the new intervention group, and thus the PFS gain associated with the new treatment is assumed to lead to an OS gain. This assumption is potentially flexible. For example, at a conservative extreme it could be assumed that a new intervention that increases PFS has no impact on OS – that is, the risk of death upon progression in the intervention group is higher than in the control group to the extent that OS is identical between the two groups. Alternatively, it could be assumed that the treatment effect is maintained into the post-progression period, with the most liberal assumption being that the treatment effect is maintained for an entire lifetime. Another option could be to assume that the treatment effect is zero after a given timepoint. These assumptions are largely arbitrary, and if such an approach was taken these should be based upon clinical data, or expert clinical knowledge including a consideration of the biological nature of the intervention and the disease itself. Even incorporating this information, assumptions are likely to remain highly debatable, and thus more complex methods that account for switching by adjusting observed survival data may be preferable.

3.1.6 Assumed equal OS for the two treatment groups

As discussed in Section 3.1.5, an extreme conservative assumption that could be made in the presence of OS data confounded by treatment switching could be that the new intervention does not confer any OS benefit, even if a PFS benefit has been demonstrated. In some cases this may be a useful analysis for decision makers, even if it is not likely to be accurate. This analysis may present a type of “worst-case” analysis for the new treatment, providing a “maximum” ICER associated with the intervention (assuming all other assumptions in the model – for example utility scores – were acceptable, and assuming that it is unlikely that the intervention would lead to a reduction in OS given a PFS gain). If this “maximum” ICER were acceptable then the decision-maker may recommend the intervention with a greater degree of confidence. However, if this analysis resulted in an ICER that was not acceptable, it would be less useful to the decision-maker without several other sensitivity analyses demonstrating the ICER for alternative OS estimates.

3.1.7 Using sequencing models

In some circumstances, it may be the case that treatment pathways are well defined, and data are available demonstrating the effectiveness of treatments given at different points in a pathway. This lends itself to a treatment sequencing economic evaluation whereby post-progression treatments are explicitly modelled as part of a treatment pathway. This represents a method for addressing the treatment switching problem because typically switching occurs after disease progression, and a sequencing model would only incorporate information from the trial of the new intervention up until the point at which the next treatment in the pathway would be administered, which would often be upon disease progression. Data on survival beyond this point would generally be taken from another source and the period confounded by treatment switching would not be used within the economic model. However, problems with the analysis would still occur if treatment switching occurred in the trial of the final-stage treatment, and often data on the effectiveness of interventions at different stages of the disease pathway are difficult to obtain.

3.2 COMPLEX METHODS

3.2.1 *Inverse Probability of Censoring Weights*

The Inverse Probability of Censoring Weights (IPCW) method represents an approach for adjusting estimates of a treatment effect in the presence of any type of informative censoring.

In the context of treatment switching, data are sorted into a panel format, with observations for individuals recorded at regular intervals through time until death or censoring. Patients are then artificially censored at the time of switch, and remaining observations are weighted based upon 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 the selection bias caused by the censoring – patients who do not switch and have similar characteristics to patients who did switch receive higher weights.

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.¹⁶ 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 for by replacing the Kaplan-Meier estimator, log-rank test, and Cox partial likelihood estimator of the hazard ratio (HR) with their IPCW versions.¹⁶

The “no unmeasured confounders” assumption represents a key limitation of the IPCW method. It cannot be tested using the observed data^{17,18} and is particularly problematic in the context of an RCT. The IPCW method represents a type of Marginal Structural Model (MSM), which were originally developed for use with observational data.^{19,20} 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, even in patients who do not switch

treatments); this can hinder the applicability of the IPCW method. Finally, the IPCW method cannot work if there are levels of any covariates which ensure (that is, the probability equals 1) treatment switching will occur.¹⁸⁻²⁰

3.2.2 Rank Preserving Structural Failure Time Model

The Rank Preserving Structural Failure Time Model (RPSFTM) was designed to address the issue of treatment non-compliance specifically in the context of RCTs. It uses a counterfactual framework to estimate the causal effect of the treatment in question,²¹ 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 which satisfies this constraint. More details on the g-estimation process are given in Appendix A. 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 (that is, 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. IV approaches use an instrument (in this case the randomised treatment group) that is predictive of the treatment to estimate causal treatment effects (see Hernan and Robins (2006)²² for further discussion on IV methods).

The RPSFTM does not rely upon the “no unmeasured confounders” assumption and identifies the treatment effect using only the randomisation of the trial, observed survival and observed treatment history. The standard one-parameter version of the model assumes 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 the randomisation of the trial means that there is only random variation between treatment groups at baseline, apart from treatment allocated – untreated survival times must be independent of the randomised treatment group.²¹ This represents the exclusion restriction assumption associated with IV methods.

The primary limitations of the standard one-parameter version of the RPSFTM involve the “common treatment effect” assumption and the randomisation (exclusion restriction) assumption. The latter should be reasonable in the context of an RCT, but the potential remains for important differences at baseline in small and in larger trials.²³ It is therefore

relevant to note that it is possible to adjust for baseline covariates within an RPSFTM analysis.²⁴ The “common treatment effect” assumption is more problematic. If patients who switch on to the experimental treatment part way through the trial receive a different treatment effect compared to patients originally randomised 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 pre-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 randomised to the experimental group. There are different ways in which the RPSFTM can be applied to a dataset – “treatment group” and “on treatment” analyses are described in Section 3.3.1. The specific approach used to apply the method needs to be justified or explored via sensitivity analysis.

3.2.3 Iterative Parameter Estimation algorithm

Branson and Whitehead (2002) extended the RPSFTM method using parametric methods, developing a novel Iterative Parameter Estimation (IPE) procedure.²⁵ 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.²⁵

The IPE procedure makes similar assumptions to the RPSFTM method – for example the randomisation 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.²⁶ Because the IPE method uses a parametric estimation procedure rather than g-estimation it may converge more quickly, but otherwise it would be expected to perform similarly, provided a suitable parametric distribution can be identified.

3.2.4 Alternative “two-stage” methods

In addition to the “standard” adjustment methods described so far, “two-stage” methods might also be considered. These methods effectively recognise that the clinical trial is randomised up until the point of disease progression, but beyond that point it essentially becomes an observational study. First a treatment effect specific to switching patients is estimated and the survival times of these patients are adjusted, subsequently allowing the treatment effect specific to experimental group patients to be estimated. Previous authors have used such an approach, making use of structural nested failure time models (SNM) with g-estimation to estimate the treatment effect in switchers.^{17,18} The SNM uses the same causal model for counterfactual survival as the RPSFTM – in fact the RPSFTM is a form of SNM. However, the distinction between an SNM and an RPSFTM is that the SNM makes the assumption of “no unmeasured confounders” rather than basing estimation on the randomisation of the trial. For this reason, the SNM has similar limitations to the IPCW.

A simplified two-stage approach that has not previously been used in an HTA context that does not rely upon g-estimation may also be considered, driven by the type of treatment switching often observed in oncology RCTs. When switching is only permitted after disease progression this timepoint can be used as a secondary “baseline” under the assumption that all patients are at a similar stage of disease at the point of disease progression. An accelerated failure time model (such as a Weibull model) that includes covariates measured at the time of progression, and including a covariate indicating treatment switch, could be fitted to the post-progression control group data to produce a reasonable estimate of the treatment effect received by patients who switched compared to control group patients who did not switch. 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. This is a simplification of the method used by Robins and Greenland¹⁷ and Yamaguchi and Ohashi¹⁸ because no attempt is made to adjust for time-dependent confounding beyond disease progression. The method therefore requires the strong assumption that there is no time-dependent confounding between the time of disease progression and the time of treatment switch. It also makes additional parametric assumptions according to which parametric accelerated failure time model is used to estimate the treatment effect in switchers.

Whilst the simple two-stage method is theoretically inferior to methods that adjust for time-dependent confounding (such as the IPCW), it has practical advantages because it does not require data to be collected on time-dependent covariates at time-points other than the secondary baseline, and hence designing a trial that satisfies the requirements of this adjustment method may be relatively straightforward. Trialists would need to ensure that switching was only permitted after a disease-related secondary baseline, and that prognostic covariate data were collected at this time-point. In addition, if switching occurs soon after the secondary baseline any time-dependent confounding associated with the lag between disease progression and treatment switching would be small.

Unlike the RPSFTM and IPE methods, the simple two-stage method does not require the “common treatment effect” assumption because the initial step of the approach involves estimating a treatment effect specifically for switchers. However, such a method may not be generalisable because it is reliant on the ability to identify a secondary baseline.

3.2.5 Using external data

In some instances it might be possible to estimate OS based upon external data, rather than relying upon confounded RCT data. External trials that incorporated the comparator treatment and that were not confounded by treatment switching may exist, or long-term registry data for the disease in question may be available. While such data sources are valuable, the use of external data may be associated with important limitations. Patient populations may differ between different trials due to inclusion criteria, and standards of care may differ if the trials were undertaken at different times and in different locations. Definitions of disease events may also differ, making it difficult to draw appropriate comparisons between trials. These issues are likely to be exacerbated further if the external data source is a registry rather than a clinical trial. If patient-level data are available from the external datasets it may be possible to adjust for differences in patient characteristics, allowing more accurate estimates of what counterfactual survival would have been in the control group of the RCT under investigation (see Section 5 for an example of this in a NICE TA). However, this requires that all important prognostic variables are available from both the novel clinical trial, and the external trial(s). In their absence, different trial populations cannot be adjusted appropriately for comparison. Finally, it may be the case that relevant external datasets do not exist, or that the patient-level data associated with these are not

available, hence using external data to adjust survival time estimates in the presence of treatment switching is unlikely to represent a generaliseable approach.

3.3 APPLICATION TO ECONOMIC EVALUATIONS

3.3.1 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 two-stage methods this involves a consideration of the plausibility of the “no unmeasured confounders” assumption. Although this assumption cannot be tested empirically, 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 to inform such judgements. The treatment switching mechanism within the trial of interest should also be explored in order to ascertain how and why treatment switching decisions were made, as this may provide information upon 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 produce reliable results. Additionally, for the simple two-stage method, no effort is made to adjust for any time-dependent confounding that occurs between the secondary baseline (for instance, disease progression) and the time of switch. Hence, the implicit assumption is that no time-dependent confounding occurs between these time-points.

For 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 have more advanced disease – receive the same benefit (per unit of time) from treatment as those in the experimental group who received the treatment from randomisation. In an attempt to relax the “common treatment effect” assumption, analysts have attempted to apply a multi-parameter version of the RPSFTM. However these have not been successful, with meaningful point estimates for causal effects difficult to determine.^{17,27,28} While some assessment of the “common treatment effect” assumption may be made using trial data (for example, by estimating the treatment effect received by switchers compared to non-switchers) such analyses are likely to be prone to time-dependent confounding and are

therefore unreliable. If patients with varying levels of disease progression were randomised into the trial of interest comparing the treatment effect in groups based upon 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 is important.

Use of the RPSFTM and IPE methods is also problematic if the comparator treatment used in the RCT is active (i.e., it prolongs survival). The RPSFTM and IPE counterfactual survival model requires that patients are either “on” or “off” at any one 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 one 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 multi-parameter 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 standard “on treatment” 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 – where patients in the experimental group are always considered to be “on” treatment and patients that switch remain “on” treatment from the time of switch until death. This analysis ignores treatment discontinuation times and estimates the effect associated with being randomised to the experimental group, rather than the effect received while taking the experimental treatment. In this sense, this approach is more similar to a standard ITT analysis of randomised groups. This approach requires there to be a common treatment effect associated with the sequence of treatments received by patients randomised to the experimental group and the sequence of treatments received by switchers after the point of switch. Any benefits associated with post-study treatments will be attributed to the experimental treatment, though similarly any benefits from post-study treatments received by control group non-switchers would be attributed to the control group. If the post-study 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 post-study 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 randomisation-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. Observational-based methods such as the IPCW and two-stage methods are not restricted in this way, but their confidence intervals may also be wide if data are relatively sparse.

3.3.2 Practical limitations

The practical limitations associated with combining treatment switching adjustment methods with economic evaluations must also be considered. Latimer (2011 and 2013) provides recommendations upon how extrapolation of survival data should be undertaken for use in economic models.^{1,26} Two main approaches are described – extrapolation using parametric models fitted independently to treatment groups; and extrapolation undertaken based upon a proportional treatment effect assumption whereby one parametric model is fitted to both treatment groups combined, with treatment group included as a covariate. Issues with both of these approaches arise when treatment switching adjustment methods are used. The RPSFTM, IPE and two-stage methods provide a counterfactual dataset that is adjusted for treatment switching, and thus either extrapolation approach can be undertaken. However, recensoring is required in order for the RPSFTM and IPE methods to avoid bias and this is also true for two-stage methods.²⁷ 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, treatment received is likely to be associated with prognosis. This means that counterfactual censoring times may be related to prognosis and may therefore be informative (see Appendix A for more details).²⁷ 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 evaluation due to the requirement to estimate the mean survival advantages associated with novel interventions.^{5-8,26,29} 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 due to 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,³⁰ after which a variety of extrapolation methods could be applied. Alternatively a variation on the proportional hazards-based extrapolation could be undertaken using the IPCW HR by fitting a parametric model to the observed experimental group 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 hazard function, from which the control group survivor function could be derived (we call this a “survivor function” approach). This may produce a degree of error because a HR is applied to an independently fitted parametric model, but this error may be minimal.

4. SIMULATION STUDIES

Simulation studies involve the simulation of data such that the “truth” is known. They allow alternative methods to be compared based upon how closely they estimate the “truth”. Several simulation studies have been undertaken in order to evaluate the performance of treatment switching adjustment methods across a wide range of scenarios.³¹⁻³³ Simulation studies are required because the “truth” must be known in order for the performance of alternative methods to be compared. Methods cannot be definitively compared through application to real world datasets, because in these datasets we do not know what would have happened in the absence of treatment switching.

Morden et al. conducted a simulation study to compare simple treatment switching adjustment approaches (such as censoring and exclusion analyses) to a limited subset of more complex methods. The more complex methods considered were:

- Adjusted Cox model³⁴
- Causal proportional hazards estimator³⁵
- Rank Preserving Structural Failure Time Model (RPSFTM)²¹
- Iterative Parameter Estimation (IPE)²⁵
- Parametric randomisation-based method³⁶

The adjusted Cox model, causal proportional hazards estimator, and parametric randomisation-based methods are not mentioned in Section 3 of this TSD for several reasons. Firstly, these methods appear sub-optimal – Morden et al. show that the adjusted Cox model and causal proportional hazards estimator were outperformed across their simulations by the RPSFTM and IPE methods. In addition, the adjusted Cox model has been shown to be highly prone to bias because it conditions on future events.³⁷ The causal proportional hazards estimator method is designed for a situation of “all-or-nothing” compliance – that is, patients who switch must switch immediately upon randomisation or not at all, which is not the case with the types of treatment switching considered in this TSD. Finally, the authors found that the parametric randomisation-based method performed very poorly and often failed to converge – hence it seems unlikely that this method should be recommended for use in HTA.³²

Morden et al. simulated independent datasets using a Weibull model in which the true treatment effect on survival was known. A baseline prognostic covariate (“good” or “poor” prognosis) which influenced the probability of switching was incorporated, but no time-dependent covariates or effects were included. It was assumed that the treatment effect was constant over time, and was equal in switchers and patients initially randomised to the experimental group – therefore, the “common treatment effect” assumption was assumed to hold. The bias, mean squared error (MSE) and coverage of each method was analysed across 16 scenarios. Weibull parameters were chosen to reflect a disease population that had a decreasing mortality rate over time, of whom 90% would have died after 3 years of follow-up (which was assumed to be the administrative censoring time). The authors tested scenarios which varied the prognosis of switching patients, the difference in survival between

prognosis groups, the probability of switching (dependent on prognosis group) and the treatment effect (HRs of 0.9 and 0.7 were tested).

The authors found that in each scenario tested bias was relatively small for the RPSFTM, although the treatment effect was slightly over-estimated, suggesting that the method was over-adjusting for treatment switching. Across the scenarios tested the IPE algorithm performed best, producing the least bias. The simple approach of censoring patients at their switching time was found to be particularly inappropriate, giving biased estimates of the true treatment effect in situations where a patient's switching pattern is strongly related to their underlying prognosis.³² Excluding patients who switched produced lower levels of bias, particularly when a low proportion of patients switched, but the bias increased as the proportion of patients that switched increased.

While the simulation study undertaken by Morden et al. is useful, it is subject to two main limitations. Firstly a “common treatment effect” was assumed in all scenarios – which satisfied the key assumption associated with randomisation-based RPSFTM and IPE methods. This assumption may be unrealistic, since patients who switch receive the experimental treatment at a more advanced stage of disease progression and may have a lower capacity to benefit. Considering this, it is important to consider how well alternative methods perform when the treatment effect is allowed to vary by group and over time. Secondly, the authors did not include any of the more complex observational-based approaches, such as IPCW or two-stage methods. It is important to consider the relative performance of these methods compared to the randomisation-based methods and the simple methods.

To address these issues Latimer et al. conducted a simulation study which incorporated a time-dependent covariate in the data generating mechanism, applied different treatment effects to switchers, and included the complex observational-based switching adjustment methods.³¹ A joint longitudinal and survival model was used 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. Scenarios 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, thus allowing the performance of each

switching adjustment method to be assessed with respect to bias, mean squared error and coverage. With respect to the RPSFTM and IPE methods in scenarios where the “common treatment effect” assumption held, the results confirmed those found by Morden et al. – that is, these methods performed very well. Also, the simple censoring and exclusion methods produced much higher levels of bias.

In addition, the authors demonstrated that the IPCW method represented a substantial improvement compared to simple methods, but produced higher bias than RPSFTM and IPE methods when the “common treatment effect” assumption held.³¹ This was likely to be due to the error associated with applying an observational-based method to a relatively small RCT dataset (with sample size of 500 patients), and was in line with the findings of other authors.³⁸ Bias associated with the IPCW method became extremely high in scenarios in which the proportion of control group patients that switched treatments increased to approximately 90%, leaving approximately 20 patients in the control group who did not switch.³¹ The authors 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 randomised to the experimental group the authors found that the RPSFTM, IPE and IPCW methods produced similar levels of bias in their estimates of the treatment effect.³¹ All produced important levels of bias, equivalent to approximately 5-10% of the treatment effect. In scenarios where the treatment effect received by switchers was approximately 25% lower than the average effect received by patients initially randomised to the experimental group the IPCW method produced lower bias than the RPSFTM and IPE methods (which often produced bias of over 10%). In these scenarios the ITT analysis often produced least bias (0-5%) if the treatment effect was relatively low (equivalent to a HR of approximately 0.75 in experimental group patients).³¹ This is logical, because in these scenarios patients who switch receive very little benefit from the experimental treatment.

Latimer et al. also tested two “two-stage” methods – a structural nested model (SNM) with g-estimation and a simple two-stage Weibull approach. The SNM performed poorly, particularly when switching proportions were very high.³¹ 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 modelling requirements. While the RPSFTM and IPE methods produced less bias than the two-stage Weibull method when the “common treatment effect” assumption held, the opposite was true when this assumption was violated. The results associated with the two-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 (and thus the scope for time-dependent confounding between the point of progression and the time of switch was limited) and prognostic covariate data were available at the time of disease progression. However, it is noteworthy that the switching mechanism simulated was similar to that observed in metastatic cancer trials, hence the good results associated with the two-stage Weibull method should not be ignored. Owing to the poor performance of the more complex adjustment methods across several scenarios, consideration of the simple two-stage method is justified in situations in which treatment switching can only occur after an identifiable secondary baseline, where switching occurs soon after that secondary baseline, where data on important prognostic factors are available at that secondary baseline. This is particularly the case in scenarios where RPSFTM, IPE and IPCW methods seem inappropriate.

Latimer et al. conducted a second simulation study, to test certain key scenario parameters that were not covered by their initial study.³³ The initial study only considered high switching proportions (approximately 65% - 95% of control group patients switched), whereas the follow-up study tested proportions ranging from approximately 10% to 95%. The follow-up study considered sample sizes varying from 300 to 500, incorporating 2:1 randomisation in favour of the experimental group, whereas the initial study considered only a sample size of 500 with 1:1 randomisation. The initial study incorporated fairly low censoring rates, – ranging between 1% and 21% whereas the follow-up study assessed higher censoring proportions – from 13% to 56%. Finally, the follow-up study tested two different

data generating models, and simulated the data using more complex hazard functions, using 2-component mixture Weibull and Gompertz models.

The authors found that generally all the switching adjustment methods produced lower levels of bias when the switching proportion was lower, as expected.³³ At very low levels of switching (less than 10%) there was a marginal increase in the bias associated with the IPCW method, reflecting the increased difficulty associated with modelling the switching process when very few patients switch. There was an increase in bias associated with all adjustment methods when the sample size was smaller, but this was marginal. The results also indicated an increase in bias associated with the adjustment methods when the censoring proportion increased, and thus higher censoring proportions combined with smaller sample sizes and higher switching proportions are likely to be problematic.

Importantly, the authors found generally lower levels of bias for all methods, even in scenarios that had similar switching proportions and treatment effect sizes (in terms of HRs) to scenarios included in their previous study. In particular, the IPCW generally produced lower levels of bias (usually less than 4%), and the impact of violating the “common treatment effect” assumption had a smaller impact on the RPSFTM and IPE methods.³³ The authors concluded that this was due mainly to the size of the acceleration factor associated with the simulated datasets. The mixture Weibull and Gompertz models used to generate the data produced survivor functions that had different shapes to those generated in the authors’ initial study – the follow-up study used more flexible survival distributions which were deemed to better reflect reality. Different shaped survivor functions meant that simulated datasets produced different average acceleration factors associated with the experimental treatment, even when the simulated average HR was similar to a corresponding scenario included in the first study. A lower acceleration factor (AF) provides less scope for the RPSFTM and IPE methods to produce bias when the common treatment effect assumption does not hold, because it is the absolute difference between the treatment effect (in terms of an AF) received by experimental group patients and the treatment effect received by switchers that causes the bias. The IPCW method also appeared to perform better when the AF was lower. Hence, it is important to assess the size of the treatment effect both in terms of a HR and in terms of an AF when investigating the likely bias associated with alternative adjustment methods. The authors found that when the acceleration factor was relatively low (less than approximately 1.8) the RPSFTM and IPE methods produced relatively low levels

of bias even when the treatment effect decrement in switchers was 20%. In addition, the IPCW method gave bias lower than the 5-10% estimated based upon the authors' first study when the acceleration factor was relatively low (less than approximately 1.8), irrespective of the hazard ratio.

Finally, Latimer et al. further investigated the two-stage adjustment method in their follow-up study – assessing a two-stage Weibull model and a two-stage Generalised Gamma model. These methods performed similarly well and often produced least bias compared to all other adjustment methods, even when the “common treatment effect” assumption held. Where this was less likely was where the “common treatment effect” assumption held and there was a lower switching proportion – in these scenarios it was more common for the randomisation-based methods to produce least bias.³³ Again, some caution must be taken with the results for the two-stage methods, because they were well suited to the simulated switching mechanism – switching could only occur soon after disease progression and so the scope for time-dependent confounding between the point of disease progression and the time of treatment switch was limited.

The results of any simulation study should be interpreted with a degree of caution. It is not possible to consider all possible scenarios that might arise in reality and therefore results may not be fully generalisable. In addition, there remains a risk that the results of simulation studies may be linked in some way to the chosen data generating process – although in their follow-up study Latimer et al. found that the performance of the adjustment methods was not affected by the data generating model.³³

5. REVIEW OF SWITCHING ADJUSTMENT METHODS USED IN NICE TAs

In this section we summarise the use of methods to adjust for treatment switching in NICE TAs. 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 – these are listed in Table 1. Of these, treatment switching occurred in pivotal clinical trials in 25 TAs (55.6%). The proportion of patients that switched differed substantially across these 25 appraisals – on two occasions less than 10% of patients

switched.^{39,40} However usually the switching proportion was much higher (for example, the proportion of control group patients that switched was higher than 40% in several TAs,⁴¹⁻⁴⁵) and would therefore be expected to have a substantial impact on overall survival time and subsequent cost-effectiveness estimates. Although the exact impact of switching on ICERs and treatment recommendations is often hard to assess from the appraisal documents, in a number of examples it was possible to determine what the ICER would have been with and without adjustments for switching. For example, in the appraisal of sunitinib for gastrointestinal stromal tumours (GIST) the ITT-based ICER of approximately £77,000 per quality adjusted life year (QALY) gained was reduced to approximately £27,000 per QALY gained when the RPSFTM was used to adjust for the switching.⁴¹ In the appraisal of imatinib for GIST the ITT-based ICER of approximately £30,000 per QALY gained was reduced to approximately £14,000 per QALY gained by excluding patients who switched from the analysis.⁴³ Therefore attempting to account for treatment switching can have an important impact on the ICER, and may have influenced the resulting recommendations made concerning the use of these treatments in the NHS.

An array of methods have been used to adjust for treatment switching in NICE TAs, demonstrating a lack of consistency between HTAs, and also a lack of clarity over which methods are appropriate. That the methods used to address treatment switching have varied across HTAs does not in itself demonstrate that methods are being used poorly, since it is likely that different methods will be appropriate in different circumstances. However, the regular use of simple censoring and exclusion techniques does show that the switching problem is being addressed sub-optimally. A break-down of the methods employed in the reviewed TAs is presented in Table 1. The methods are categorised into “simple” and “more complex” approaches, defined by the complexity of the statistical approach taken to adjust specifically for treatment switching – reflecting the categorisation used in Section 3 of this TSD.

Table 1: Methods used to account for switching in NICE technology appraisals (2000 – 2009)

Method	TAs which use method
<i>“Simple” methods</i>	
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)
Modelled based on PFS, not OS	2 (TAs 6, 33)
Used sequencing models	2 (TAs 93, 176)
Applied the same risk of death upon disease progression	1 (TA 118)
Assumed equal OS for the two treatment groups	1 (TA 119)
<i>More complex methods</i>	
Rank preserving structural failure time model (RPSFTM)	1 (TA 179)
Adjusted survival estimates using a case-mix approach	1 (TA 34)
Used external data	1 (TA 171)

Note: The numbers in this Table do not sum to 25 because in 6 TAs more than one method was used.

In the TAs in which methods were used to adjust for treatment switching, censoring and exclusion approaches were most common (used in 11 of the 25 TAs (44%)); these approaches are clearly associated with selection bias.^{14,15} In 7 (28%) TAs treatment switching was not addressed at all. The simple approach of including the costs of switching treatments generally does not meet the requirements of the economic evaluation decision problem, while modelling based upon PFS rather than OS, applying the same risk of death upon disease progression in each treatment group, or assuming equal OS for the two treatment groups makes no use of the data collected on the treatment effect on post progression survival. Sequencing models, whereby post-progression treatments are explicitly modelled as part of a treatment pathway, were occasionally used. These may avoid the issues created by treatment switching after disease progression if unconfounded data for each treatment in the sequence are available – however this is often not the case and in the two TAs that took this approach the final treatment sequence modelled remained potentially confounded by treatment switching.^{46,47}

Only one TA (sunitinib for GIST, TA179) used a recognised complex switching adjustment method (Robins and Tsiatis's RPSFTM²¹).⁴¹ In one TA (trastuzumab for breast cancer, TA34) a case-mix approach which appeared similar to an IPCW method was used to adjust for treatment switching,⁴⁸ however very few details on this were presented in the appraisal documents.

Recently there has been a tendency towards the use of more complex treatment switching adjustment methods such as RPSFTM and IPCW. For example, in two NICE appraisals completed since we completed our review (pazopanib for the first-line treatment of metastatic RCC (TA215) and everolimus for the second-line treatment of advanced RCC (TA219)) both RPSFTM and IPCW methods were used.^{49,50} However, there remains evidence of uncertainty around which methods are appropriate for adjusting for 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 in the appraisal documents.^{49,50} Hence, while the RPSFTM method appeared to be preferred in these appraisals, there was no evidence that the advantages and disadvantages associated with each method had been fully taken into account and, from the appraisal documents, it is not clear that the most appropriate switching adjustment method was identified.

5.1 EXTERNAL DATA

The focus in this technical support document is upon statistical methods that may be used to adjust observed survival data in the presence of treatment switching. However, in one TA (lenalidomide for multiple myeloma, TA171), a different approach was taken: external data were used in an attempt to adjust for treatment switching.⁴⁵ Patient-level data from two external trial datasets were used in order to estimate what post-progression survival would have been in the novel clinical trial had treatment switching not occurred. In the key novel trial approximately 50% of control group patients switched onto lenalidomide, with 75% of that crossover occurring after disease progression.⁵¹ To address this, the manufacturer used patient-level data from previous trials that included similar (but not identical) control group

arms that were not confounded by switching. The manufacturer provided analyses to demonstrate that the OS that could be expected for the control group treatment (dexamethasone) used in their novel lenalidomide trial was similar to that observed in the external trials (which used dexamethasone as well as some other standard treatments as control).⁴⁵ In addition, the manufacturer produced an analysis to demonstrate that there was no evidence of an OS improvement over time.⁴⁵ This was important because the external trial datasets were dated, with patients enrolled between 1980 and 1997. Based upon these analyses, the manufacturer rationalised the use of the external trial datasets for inferring what control group survival in the novel lenalidomide trial would have been, had treatment switching not occurred.

The manufacturer fitted parametric survival models to the external trial data in order to derive an equation for OS that included a range of patient characteristic variables.⁴⁵ The values of these variables were then set to reflect the patient characteristics observed in the lenalidomide trial, and hence survival times that would have been observed in the external trial had the patient characteristics in the control arm matched those in the novel lenalidomide trial were estimated. The manufacturer did not use this estimate of OS directly in the economic model, because PFS and post-progression survival (PPS) were modelled as distinct states, with PFS estimated based only on the novel lenalidomide trial (this in itself is questionable, since 25% of switching occurred before disease progression). In the economic model the manufacturer used a “calibration factor” applied to PPS such that the median OS estimated from the external trial dataset adjusted for the lenalidomide trial patient characteristics equalled the median OS estimated by the model, as a function of PFS plus PPS.

The Assessment Group noted some problems with the manufacturer’s analysis.⁴⁵ Firstly, they noted that mean OS rather than median OS should have been used to calibrate the estimated OS in the control arm of the lenalidomide trial with the external data. A second problem highlighted by the Assessment Group was that there were likely to be important patient characteristics not reported in both the novel lenalidomide trials and the external trials which could not be included in the OS equations. Hence it may not have been possible to fully adjust the external trial survival estimates to reflect the lenalidomide trial patient population. The analysis is essentially reliant on a “no unmeasured confounders” assumption, and the lack of analysis to identify any important variables missing from either the lenalidomide or external trial datasets represented an important oversight on the part of

the manufacturer. Finally, the Assessment Group noted that alternative data sources suggested improvements in survival in the relevant patient group between 1995 and 2006, thus suggesting the dated external trials may indeed represent an underestimate of present-day control group OS.⁴⁵

An additional issue which was not mentioned by the Assessment Group but was discussed by the Appraisal Committee surrounded the clinical validity of the manufacturer's analysis.⁵¹ There were two lenalidomide trials relevant to the appraisal, and the application of the manufacturer's analyses to these trials led to control group OS estimates that were approximately half those observed in the trials themselves. These details were marked as "commercial-in-confidence" in the TA documents, but were reported in a subsequent published paper.⁵² Therefore, based upon the manufacturer's analysis, the impact of approximately 50% of control group patients switching onto lenalidomide was to cause the mean OS for the control group as a whole to approximately double. For this to be the case, the experimental treatment would have to more than double life expectancy for switchers. In the key lenalidomide clinical trial the gain in PFS for lenalidomide was large: 13.4 months compared to 4.6 months in the control arm (2.9 times longer for lenalidomide). Therefore a similar relative effect on OS could potentially lead to the OS estimates derived by the manufacturer. However, this would assume that the relative effect of lenalidomide on OS is the same (if not higher) than for PFS, and that receiving lenalidomide after disease progression leads to the same (if not higher) impact on OS as is the case when it is given before disease progression. The Appraisal Committee noted that the manufacturer's approach led to an improvement in OS predicted by the economic model which was out of proportion given the improvement seen in PFS.⁵¹ Despite these issues, the deliberations of the Appraisal Committee regarding TA171 demonstrated openness to the use of external data in the presence of treatment switching. Such an approach is not generalizable though, because often suitable external datasets will not be available, as mentioned in Section 3.2.5.

5.2 REVIEW CONCLUSIONS

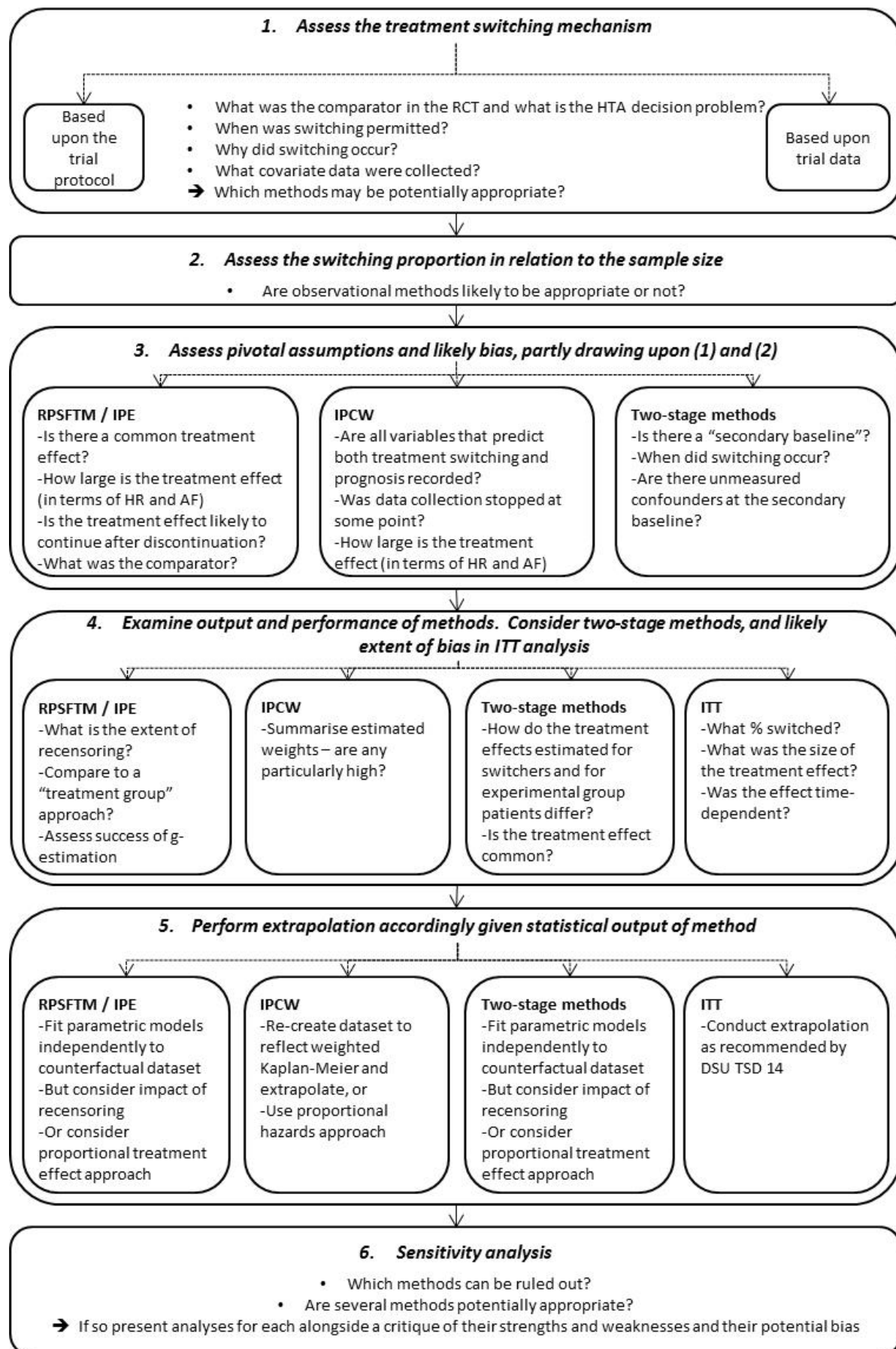
It is clear that alternative complex adjustment methods make very different assumptions and work in very different ways, hence they are likely to produce different results. This has been demonstrated in HTA; in the NICE appraisal of pazopanib for the first-line treatment of metastatic renal cell carcinoma (RCC) the IPCW method produced an ICER of approximately

£49,000 per QALY gained, whereas the RPSFTM method produced an ICER of approximately £33,000 per QALY gained.⁴⁹ While there has been a trend towards using more complex methods in HTA these remain poorly discussed and inadequately justified. Two-stage methods appear to be potentially useful methods that have not previously been used in HTA.

6. METHODOLOGICAL AND PROCESS GUIDANCE

Based upon a 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 upon 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 (see Figure 2). It is important to note that these recommendations refer specifically to methods that adjust observed data in the presence of treatment switching; they do not incorporate methods such as the use of external datasets. However, when treatment switching arises, the possibility and practicality of using external data in order to estimate counterfactual survival times should be considered. In addition, it should be noted that switching adjustment methods may be used in tandem with external data – switching adjustment concerns the events observed in the trial period, whereas economic models are often required to extrapolate into the future. First data confounded by switching must be adjusted, and then the counterfactual data must be extrapolated – this is dealt with briefly in Step (5) of Figure 2, but is discussed in more detail in TSD 14,¹ which states that external validity and clinical plausibility is of the utmost importance in survival projections. Hence an investigation of relevant external datasets is likely to be useful whether or not treatment switching occurs.

Figure 2: Treatment switching analysis framework



Step (1) involves assessing the treatment switching mechanism and considering this in relation to the decision problem faced in the technology appraisal. This should demonstrate whether and which adjustment methods are potentially applicable and relevant. For instance, it may become apparent whether data on relevant switching indicators were collected (if they were not, the IPCW method is unlikely to be appropriate), or whether the comparator included in the RCT was relevant for the decision problem. The time at which patients became able to switch treatments is also important to determine, as this helps identify whether two-stage methods are likely to be applicable (these will only be appropriate if switching is only permitted after a certain disease-related time-point).

For Step (2), the proportion of patients switching treatment 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, since sample sizes are rarely larger than the size of 500 (250 in each arm) tested in Latimer et al.'s initial simulation study.³¹ It is likely that the sample size would need to be substantially greater than 500 in order for the IPCW to produce unbiased results when the proportion of patients that switch is as high as 90%. Further, problems may arise even with lower switching proportions. For instance, if only 50% of control group patients switch, but this represents 90% of those patients who experienced disease progression (and thus became eligible to switch), the IPCW method will be prone to bias: it is the switching proportion in patients who became eligible to switch that is of primary importance. A similar situation could occur if switching was only permitted in specific patients – for instance, those who had previously responded to treatment, or those with a specific biomarker present. Randomisation-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 further assessing the pivotal assumptions of each of the adjustment methods in order 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 randomised group included as a covariate and a switching indicator variable may be used, but the potential bias associated with these should be recognised. Depending upon 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 randomised into 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 the assumption should be routinely considered. It is important to harness what is known by a variety of scientists and clinicians about the impact of patient characteristics and disease progression on the effects of the drug being studied. If these analyses suggest that the “common treatment effect” assumption holds an RPSFTM or IPE approach should be used. An IPCW approach may also produce low bias, but this is less certain. For RPSFTM, IPE and IPCW methods it is important to consider the size of the treatment effect both in terms of a hazard ratio (HR) and an acceleration factor (AF).

When using RPSFTM or IPE methods the duration of the treatment effect (i.e. whether it is likely to be maintained to any extent after treatment discontinuation) must be considered. If it is likely that the treatment effect may be maintained beyond treatment discontinuation a “treatment group” application (or the use of a lagged treatment effect) might be considered. The decision of whether to take the standard “on treatment” approach or the “treatment group” approach should be justified based upon the economic evaluation decision problem, clinical opinion, biological plausibility and data availability. It is likely to be appropriate to present each analysis, in order that the sensitivity of survival estimates and cost-effectiveness results to these can be shown. Clinical expert opinion on whether treatment advantages are likely to cease, continue, or be reversed after treatment discontinuation may be important in justifying the chosen approach. The comparator included in the RCT (i.e. whether active or not) must also be considered. If the comparator is active the RPSFTM and IPE methods may not be appropriate, although a “treatment group” approach may be justified based upon assumptions made about the treatment pathways observed in the trial.

It is important to note that a standard “on treatment” application of the RPSFTM or IPE methods provides an estimate of the treatment effect associated with *full* treatment with the experimental intervention – that is, it represents the treatment effect that would have been

observed if all patients in the experimental group received the experimental treatment throughout the trial (with no discontinuation) compared to zero treatment in the control group. Usually this is not an appropriate treatment effect for the economic evaluation, because treatment discontinuation observed in the clinical trial is likely to reflect discontinuation that would occur in the real world. Therefore, although it is valid to estimate untreated control group survival times using an “on treatment” approach, under the assumption that the treatment effect disappears upon discontinuation, these survival times should be compared to the *observed* experimental group survival times in order to provide a valid adjusted estimate of the treatment effect.

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 switch, 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 (for example, 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 two-stage methods the existence of an appropriate secondary baseline (such as disease progression) is pivotal. These will only exist if there is a timepoint before which treatment switching could not occur. If such a time-point exists two-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. Whilst simulation studies have provided support for the use of two-stage methods,^{32,33,53} it should be recognised that further research on these methods – particularly on their sensitivity to departures from their assumptions (such as the proximity of switch to the secondary baseline, and the “no unmeasured confounders” assumption) would be valuable.

After applying the switching adjustment methods Step (4) involves a review of the output of the methods in order to help identify whether the methods are likely to have performed well. For RPSFTM, IPE and two-stage methods this includes a consideration of the degree of recensoring, and possibly a comparison of standard RPSFTM and IPE results to the results obtained 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 RPSFTM and IPE methods produce treatment effects that result in equal counterfactual survival times between randomised 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 two-stage methods may be used to help determine the appropriateness of other methods – for instance, if the two-stage methods produce estimates of the treatment effect in the switching patients that are (not) similar to the effect estimated for patients randomised to the experimental group the RPSFTM / IPE methods may (not) be appropriate.

In tandem with a consideration of complex switching adjustment methods, the results of a standard ITT analysis should be considered: if other methods are likely to have performed poorly the ITT analysis may provide least bias. If the treatment effect is small (with a HR of approximately 0.75-1.00 in the experimental group, based upon the simulation study by Latimer et al.³¹) 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 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 IPCW and RPSFTM / IPE methods unless the treatment effect is high (equivalent to a HR of approximately 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 least bias.

After adjustment methods have been assessed based upon their theoretical and practical suitability as well as their performance in Steps 1-4, Step (5) addresses combining the

adjustment methods with an extrapolation approach (if required). This is based upon the statistical output of the applied adjustment method. For the RPSFTM, IPE and two-stage methods an analysis investigating the impact of recensoring on the tail of the counterfactual Kaplan-Meier curve should be undertaken to identify whether recensoring is likely to lead to inappropriate extrapolations. A “survivor function” approach whereby the adjusted treatment effect is applied to an extrapolation of uncensored experimental group survival times may be preferable. However the choice of extrapolation method should follow the advice offered by NICE DSU Technical Support Document 14 where possible.^{1,26} For IPCW appropriate methods should be used to recreate a dataset to reflect the weighted Kaplan-Meier if a proportional hazards approach to extrapolation is not to be taken.

Finally, when preliminary analysis of trial data suggests that the choice of preferable adjustment method is unclear, sensitivity analysis should be undertaken to demonstrate the uncertainty associated with the methodology used.

7. DISCUSSION

Treatment switching adjustment methods have often been used poorly and have been inadequately described in economic evaluations. The review of NICE TAs presented in Section 5 of this TSD demonstrates that while some potentially appropriate methods have been used, more often simple methods that are highly prone to bias have been relied upon. 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 – thus failing to consider their key limitations.^{49,50} This is important because the application of switching adjustment methods within an economic model often drastically alters the estimated incremental cost effectiveness ratio. The analysis framework presented in Section 6 aims to reduce the use of inappropriate and inconsistent methods, by promoting a rigorous procedure for identifying and justifying appropriate switching adjustment methods.

Because the RPSFTM/IPE and IPCW methods work in very different ways and make very different assumptions, one individual method is unlikely to always be better than the other. Trial and switching characteristics must be considered on a case-by-case basis in order to assess which switching adjustment method is likely to be most appropriate. The IPCW

method has observational data origins and its reliance on the “no unmeasured confounders” assumption represents a very important limitation which may be difficult to justify in an RCT setting. 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 two-stage methods should be considered, particularly in circumstances in which 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, this is at the cost of the potentially even stronger assumption that there is no time-dependent confounding between the secondary baseline and the time of switch. Where switching occurs soon after the secondary baseline the scope for such time-dependent confounding is limited, but this is not the case if switching happens substantially after the secondary baseline.

While the analysis framework presented in Section 6 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 (for example, approximately 90% in a control group sample size in the region of 250 subjects) 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 two-stage methods also to become prone to error and bias; although this was not demonstrated in Latimer et al.’s simulation studies,^{31,33} 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.

It is clear 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 in order 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 randomised to the experimental treatment an ITT analysis may result in relatively low bias.

This TSD focuses upon adjusting survival time estimates in the presence of treatment switching from the control treatment onto the experimental treatment. In some circumstances it may be desirable to also adjust for switching from the experimental treatment onto the control treatment, or for switching onto other alternative therapies – although often such switches may represent realistic treatment pathways that do not require adjustment within an economic evaluation context. RPSFTM and IPE methods are designed to cope with treatment switching in either direction (provided the control treatment is placebo, or non-active), but are not suitable when switching is to a third treatment. In such circumstances a multi-parameter RPSFTM would be required, but these have been shown to perform poorly in practice.^{17,27,28} Theoretically IPCW and two-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. Alternatively, when switching from the control group to the experimental treatment is followed by a switch to a post-study treatment and adjusting for both of these switches is required, combining RPSFTM (to adjust for the initial switch) and IPCW (to adjust for the post-study treatment) may be considered.

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 costs of treatments that were switched to, or only considering quality of life scores in non-switchers, we are unaware of attempts to adjust for the effects of switching on these outcomes in HTA. The problem may not be as serious as for survival estimates – quality of life scores are often based upon health states rather than treatment group, and direct and indirect costs are often based upon assumptions or external sources – but further research in these areas would be valuable. 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.⁹⁻¹² Adjusting for these outcomes is not discussed in detail in this TSD.

Finally, it is important to recognise that this TSD focusses upon 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 (for example, external trials not confounded by switching, or registry data) will not be available, but where it is methods to formally synthesise data would have value. This is particularly important because the statistical adjustment methods focussed upon in this TSD 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, this TSD only considers 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, is ongoing.¹³ Also, we only briefly discuss combining extrapolation methods with switching adjustment methods in this TSD – further research in this area would be beneficial.

8. CONCLUSIONS

It is clear that treatment switching is an important factor in a substantial proportion of HTAs, particularly in the oncology setting. This TSD 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.

9. REFERENCES

1. Latimer NR. NICE DSU Technical Support Document 14: Survival analysis for economic evaluations alongside clinical trials - extrapolation with patient-level data. 2011; available from <http://www.nicedsu.org.uk/NICE%20DSU%20TSD%20Survival%20analysis.updated%20March%202013.pdf>
2. Latimer, N.R., Abrams, K.R., Lambert, P.C., Crowther, M.J., Wailoo, A.J., Morden, J.P. et al. Adjusting Survival Time Estimates to Account for Treatment Switching in Randomized Controlled Trials - an Economic Evaluation Context: Methods, Limitations, and Recommendations. *Medical Decision Making* 2014.
3. Center for Drug Evaluation and Research (CDER). US Department of Health and Human Services Food and Drug Administration. Guidance for Industry: Clinical trial endpoints for the approval of cancer drugs and biologics. 2007; available from <http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/ucm071590.pdf>
4. Committee for Medicinal Products for Human Use (CHMP). 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. European Medicines Agency, editor. 2006.
5. National Institute for health and Clinical Excellence. Guide to the methods of technology appraisal. 2013; available from <http://publications.nice.org.uk/guide-to-the-methods-of-technology-appraisal-2013-pmg9> (accessed July 2013).
6. Briggs, A., Claxton, K., Sculpher, M. Decision modelling for health economic evaluation. Oxford University Press Inc., New York; 2006.
7. Gold, M.R., Siegel, J.E., Russell, L.B., Weinstein, M.C. Cost-effectiveness in health and medicine. Oxford University Press, USA, 1996.
8. Canadian Agency for Drugs and Technologies in Health. Guidelines for the economic evaluation of health technologies. 2006. Canada, 3rd edition.
9. Goetghebeur, E., Lapp, K. The effect of treatment compliance in a placebo-controlled trial: Regression with unpaired data. *Applied Statistics* 1997; 46:351-364.
10. Fischer-Lapp, K., Goetghebeur, E. Practical properties of some structural mean analyses of the effect of compliance in randomized trials. *Controlled Clinical Trials* 1999; 20:531-546.
11. Robins, J.M. Correcting for non-compliance in randomized trials using structural nested mean models. *Communications in Statistics-Theory and Methods* 1994; 23:2379-2412.
12. White, I.R. Uses and limitations of randomization-based efficacy estimators. *Statistical Methods in Medical Research* 2005; 14(4):327-347.

13. Boucher, R.H., Abrams, K.R., Crowther, M.J., Lambert, P.C., Morden, J.P., Wailoo, A.J. et al. PRM201: Adjusting for treatment switching in clinical trials when only summary data are available - An evaluation of potential methods. *Value In Health* 2013; 16, A323 - A636.
14. Lee, Y., Ellenberg, J.H., Hirtz, D.G., Nelson, K.B. Analysis of clinical trials by treatment actually received: is it really an option? *Statistics in Medicine* 1991; 10(10):1595-1605.
15. Horwitz, R.I., Horwitz, S.M. Adherence to treatment and health outcomes. *Archives of Internal Medicine* 1993; 153(16):1863.
16. Robins, J.M., Finkelstein, D.M. 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):779-788.
17. Robins, J.M., Greenland, S. Adjusting for differential rates of prophylaxis therapy for PCP in high-versus low-dose AZT treatment arms in an AIDS randomized trial. *Journal of the American Statistical Association* 1994; 89(427):737-749.
18. 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. *Statistics in Medicine* 2004; 23(13):1991-2003.
19. Hernan, M.A., Brumback, B., Robins, J.M. Marginal structural models to estimate the joint causal effect of nonrandomized treatments. *Journal of the American Statistical Association* 2001; 96(454):440-448.
20. Robins, J.M. Marginal structural models versus structural nested models as tools for causal inference. *Statistical models in epidemiology, the environment, and clinical trials*. Springer; 2000; 95-133.
21. Robins, J.M., Tsiatis, A.A. Correcting for non-compliance in randomized trials using rank preserving structural failure time models. *Communications in Statistics-Theory and Methods* 1991; 20(8):2609-2631.
22. Hernan, M.A., Robins, J.M. Instruments for causal inference: An Epidemiologist's dream? *Epidemiology* 2006; 17:360-372.
23. Senn S.J. Covariate imbalance and random allocation in clinical trials. *Statistics in Medicine* 1989; 8(4):467-475.
24. Hampson L.V., Metcalfe C. Incorporating prognostic factors into causal estimators: a comparison of methods for randomised controlled trials with a time-to-event outcome. *Statistics in Medicine* 2012; 31(26):3073-3088.
25. Branson, M., Whitehead, J. Estimating a treatment effect in survival studies in which patients switch treatment. *Statistics in Medicine* 2002; 21(17):2449-2463.

26. Latimer, N.R. Survival Analysis for Economic Evaluations Alongside Clinical Trials: Extrapolation with Patient-Level Data. Inconsistencies, Limitations, and a Practical Guide. *Medical Decision Making* 2013.
27. White, I.R., Babiker, A.G., Walker, S., Darbyshire, J.H. Randomization-based methods for correcting for treatment changes: examples from the Concorde trial. *Statistics in Medicine* 1999; 18(19):2617-2634.
28. 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. *Statistics in Medicine* 2004; 23(13):2005-2022.
29. Guyot, P., Welton, N.J., Ouwens, M.J., Ades, A.E. Survival time outcomes in randomized, controlled trials and meta-analyses: the parallel universes of efficacy and cost-effectiveness. *Value In Health* 2011; 14(5):640-646.
30. Guyot, P., Ades, A.E., Ouwens, M.J., Welton, N.J. Enhanced secondary analysis of survival data: reconstructing the data from published Kaplan-Meier survival curves. *BMC Medical Research Methodology* 2012; 12(1):9.
31. Latimer, N., Abrams, K., Lambert, P., Crowther, M.J., Wailoo, A., Morden, J.P. et al. Adjusting for treatment switching in randomised controlled trials – a simulation study. *University of Sheffield Health Economics and Decision Science Discussion Paper No.13/06* 2013. 2013.
32. Morden, J., Lambert, P., Latimer, N., Abrams, K., Wailoo, A. Assessing methods for dealing with treatment switching in randomised controlled trials: a simulation study. *BMC Medical Research Methodology* 2011; 11(1):4.
33. Latimer, N.R., Abrams, K.R., Lambert, P.C., Crowther, M.J., Morden, J.P. Assessing methods for dealing with treatment crossover in clinical trials: A follow-up simulation study. *University of Sheffield Health Economics and Decision Science Discussion Paper No 14/01* 2014.
34. Law M.G., Kaldor J.M. Survival analyses of randomized clinical trials adjusted for patients who switch treatments. *Statistics in Medicine* 1996; 15(19):2069-2076.
35. Loeys T., Goetghebeur E. A causal proportional hazards estimator for the effect of treatment actually received in a randomized trial with all-or-nothing compliance. *Biometrics* 2003; 59(1):100-105.
36. Walker, A.S., White, I.R., Babiker, A.G. Parametric randomisation-based methods for correcting for treatment changes in the assessment of the causal effect of treatment. *Statistics in Medicine* 2004; 23(4):571-590.
37. White, I.R. Survival analysis of randomized trials with treatment switching. *Statistics in Medicine* 1997; 16(22):2619-2620.
38. Howe, C.J., Cole, S.R., Chmiel, J.S., Muñoz, A. Limitation of inverse probability-of-censoring weights in estimating survival in the presence of strong selection bias. *American Journal of Epidemiology* 2011; 173(5):569-577.

39. Vermorken, J.B., Mesia, R., Rivera, F., Remenar, E., Kawecki, A., Rottey, S. et al. Platinum-based chemotherapy plus cetuximab in head and neck cancer. *New England Journal of Medicine* 2008; 359(11):1116-1127.
40. Roche Products Ltd. Achieving clinical excellence in the treatment of relapsed non-small cell lung cancer, Tarceva (erlotinib). 2006. NICE STA submission.
41. 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. *Exeter, UK: Peninsula Technology Assessment Group (PenTAG)* 2009.
42. Lewis, R., Bagnall, A.M., Forbes, C., Shirran, E., Duffy, S., Kleijnen, J. et al. A rapid and systematic review of the clinical effectiveness and cost-effectiveness of trastuzumab for breast cancer. *Technology Assessment Report Commissioned by the NHS R&D HTA Programme on Behalf of the National Institute for Clinical Excellence* 2001.
43. National Institute for health and Clinical Excellence. Final Appraisal Determination: Imatinib for the treatment of unresectable and/or metastatic gastro-intestinal stromal tumours, TA86. 2004. London, NICE.
44. Janssen-Cilag Ltd. STA submission to NICE: Velcade (Bortezomib) for the treatment of multiple myeloma patients at first relapse. 2006.
45. 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. *Submission to NICE As Part of STA Program* 2008.
46. Hind, D., Tappenden, P., Tumor, I., Eggington, S., Sutcliffe, P., Ryan, A. Technology assessment report commissioned by the HTA Programme on behalf of the National Institute for Clinical Excellence: 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. Produced by The School of Health and Related Research, University of Sheffield. January 2005. 2010.
47. Merck Serono Ltd. Single Technology Appraisal Submission: Erbitux (cetuximab) for the first-line treatment of metastatic colorectal cancer. 2008.
48. National Institute for health and Clinical Excellence. Guidance on the use of trastuzumab for the treatment of advanced breast cancer, NICE Technology Appraisal Guidance No 34. 2002. London, NICE.
49. National Institute for health and Clinical Excellence. Pazopanib for the first line treatment of metastatic renal cell carcinoma, TA 215. 2011. London. NICE.
50. National Institute for health and Clinical Excellence. Everolimus for the second-line treatment of advanced renal cell carcinoma, TA219. 2011. London, NICE.

51. National Institute for Health and Clinical Excellence. Final Appraisal Determination: Lenalidomide for the treatment of multiple myeloma in people who have received at least one prior therapy. 2009; TA171.
52. Ishak, K.J., Caro, J.J., Drayson, M.T., Dimopoulos, M., Weber, D., Augustson, B. et al. Adjusting for patient crossover in clinical trials using external data: a case study of lenalidomide for advanced multiple myeloma. *Value In Health* 2011; 14(5):672-678.
53. Latimer, N.R., Abrams, K.R., Lambert, P.C., Crowther, M.J., Wailoo, A.J., Morden, J.P. et al. Adjusting for treatment switching in randomised controlled trials – a simulation study. *University of Sheffield Health Economics and Decision Science Discussion Paper No 13/06* 2013.
54. Robins, J.M. Structural nested failure time models. *Encyclopedia of Biostatistics* 1998.
55. Mark, S.D., Robins, J.M. A method for the analysis of randomized trials with compliance information: An application to the multiple risk factor intervention trial. *Controlled Clinical Trials* 1993; 14(2):79-97.

APPENDIX A: COMPLEX SWITCHING ADJUSTMENT METHODS

IPCW

Robins and Finkelstein (2000) recommend using “stabilised” inverse probability of censoring weights, as these are shown to be more efficient.¹⁶ Unstabilised weights are simply the inverse of the conditional probability of having remained uncensored until time t conditional on baseline and time-dependent covariates, whereas stabilised weights are the conditional probability of having remained uncensored until time t given baseline covariates, divided by the conditional probability of having remained uncensored until time t given baseline and time-dependent covariates. The stabilised weight will be equal to 1 for all t if the history of the included prognostic factors for failure do not impact upon the hazard of censoring at t – thus there would be no informative censoring and treatment switching would be random.¹⁶ Formally, the stabilised weights applied to each individual for time interval (t) , as specified by Hernan *et al.* are:¹⁹

$$\hat{W}(t) = \prod_{k=0}^t \frac{\Pr[C(k)=0|\bar{C}(k-1)=0,\bar{A}(k-1),V,T>k]}{\Pr[C(k)=0|\bar{C}(k-1)=0,\bar{A}(k-1),\bar{L}(k),T>k]} \quad [A1]$$

where $C(k)$ is an indicator function demonstrating whether or not informative censoring (switching) had occurred at the end of interval k , and $\bar{C}(k-1)$ denotes censoring history up to the end of the previous interval $(k-1)$. $\bar{A}(k-1)$ denotes an individual’s treatment history up until the end of the previous interval $(k-1)$, and V is an array of an individual’s baseline covariates. $\bar{L}(k)$ denotes the history of an individual’s time-dependent covariates measured at or prior to the beginning of interval k , and includes V . Hence the numerator of (2) represents the probability of an individual remaining uncensored (i.e. not having switched) at the end of interval k given that that individual was uncensored at the end of the previous interval $(k-1)$, conditional on baseline characteristics and past treatment history. The denominator represents that same probability conditional on baseline characteristics, time-dependent characteristics and past treatment history. When the cause of informative censoring is treatment switching, past treatment history is removed from the model because as soon as switching occurs the individual is censored.

The IPCW adjusted Cox hazard ratio (HR) can be estimated by fitting a time-dependent Cox model to a dataset in which switching patients are artificially censored. The model includes baseline covariates and uses the time-varying stabilised weights for each patient and each time interval. Robust variance estimators or bootstrapping should be used to estimate confidence intervals.^{19,20}

RPSFTM

An accelerated failure time counterfactual survival model such as that presented by Robins (1998) is used:⁵⁴

$$U_i = \int_0^{T_i} \exp[\psi A_i(t)] dt \quad [A2]$$

where U is the counterfactual survival time for each patient, which is a known function of observed survival time (T), observed treatment ($A_i(t)$, where $A_i(t)$ is a binary time-dependent variable equal to 1 or 0 over time), and the unknown treatment effect parameter ψ .

Counterfactual survival time is a sum of observed time spent on treatment and observed time spent off treatment, where time spent on treatment is multiplied by the factor $\exp(\psi)$. g-estimation involves testing a series of potential values for ψ , and the value of the treatment effect (ψ_0) is estimated as the value of ψ for which counterfactual survival is independent of randomised groups. Within the g-estimation process a log-rank or Wilcoxon test can be used for the RPSFTM g-test in a non-parametric setting, testing the hypothesis that the baseline survival curves are identical in the two treatment groups, or a Wald test could be used for parametric models.⁵⁵ The log-rank test is conventional, and weights each risk set equally. It may be optimal if there are proportional hazards. However, if hazards are not proportional over time an alternative test – such as the Wilcoxon, which weights by the number in each risk set – may be preferable. The point estimate of ψ is that for which the test (z) statistic equals zero. Because the RPSFTM is a randomisation-based efficacy estimator (RBEE) the p-value from the ITT analysis is maintained.²⁷

White *et al.* demonstrate that censoring is problematic for the RPSFTM.²⁷ 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, treatment received is likely to be associated with prognosis. In turn, this means that the censoring of counterfactual survival times may depend on prognostic factors and therefore be informative.²⁷ Bias associated with this can be avoided by recensoring counterfactual survival times at the earliest possible censoring time given the treatment effect ψ .²⁷ Thus for each patient in treatment groups at risk of switching the recensored censoring time is the minimum of the observed administrative censoring time (C_i) and the product $\exp(\psi)C_i$. If the patient experienced an event, but the recensoring time is less than the event time, that patient has their survival time recensored and their event is no longer observed.

IPE ALGORITHM

This method uses the same accelerated failure time model as the RPSFTM, but a parametric failure time model is fitted to the original, unadjusted ITT data to obtain an initial estimate of ψ . The observed failure times of switching patients are then re-estimated using $\exp(\psi)$ and the counterfactual survival time model presented in equation [A2], and the treatment groups are then compared again using a parametric failure time model. This will give an updated estimate of ψ , and the process of re-estimating the observed survival times of switching patients is repeated. This iterative process is continued until the new estimate for $\exp(\psi)$ is very close to the previous estimate (the authors suggest within 10^{-5} of the previous estimate but offer no particular rationale for this), at which point the process is said to have converged.²⁵ Bootstrapping is recommended to obtain standard errors and confidence intervals for the treatment effect.²⁵

Table A 1: NICE Technology Appraisals (TAs) included in the review

TA Number	Title	Disease Stage	Date Issued
TA3	Ovarian cancer - taxanes (replaced by TA55)	Advanced	May 2000
TA6	Breast cancer - taxanes (replaced by TA30)	Advanced	Jun 2000
TA23	Brain cancer - temozolomide	Advanced	Apr 2001
TA25	Pancreatic cancer - gemcitabine	Advanced / Metastatic	May 2001
TA26	Lung cancer - docetaxel, paclitaxel, gemcitabine and vinorelbine (updated by and incorporated into CG24 Lung cancer)	Advanced / Metastatic	Jun 2001
TA28	Ovarian cancer - topotecan (replaced by TA91)	Advanced	Jul 2001
TA29	Leukaemia (lymphocytic) - fludarabine (replaced by TA119)	Advanced	Sep 2001
TA30	Breast cancer - taxanes (review)(replaced by CG81)	Advanced	Sep 2001
TA34	Breast cancer - trastuzumab	Metastatic	Mar 2002
TA33	Colorectal cancer (advanced) - irinotecan, oxaliplatin & raltitrexed (replaced by TA93)	Advanced	Mar 2002
TA37	Lymphoma (follicular non-Hodgkin's) - rituximab (replaced by TA137)	Advanced / Metastatic	Mar 2002
TA45	Ovarian cancer (advanced) - pegylated liposomal doxorubicin hydrochloride (replaced by TA91)	Advanced	Jul 2002
TA50	Leukaemia (chronic myeloid) - imatinib (replaced by TA70)	All stages	Oct 2002
TA54	Breast cancer - vinorelbine (replaced by CG81)	Advanced / Metastatic	Dec 2002
TA55	Ovarian cancer - paclitaxel (review)	Advanced	Jan 2003
TA62	Breast cancer - capecitabine (replaced by CG81)	Advanced / Metastatic	May 2003
TA61	Colorectal cancer - capecitabine and tegafur uracil	Metastatic	May 2003
TA65	Non-Hodgkin's lymphoma - rituximab	Advanced / Metastatic	Sep 2003
TA70	Leukaemia (chronic myeloid) - imatinib	All stages	Oct 2003
TA86	Gastro-intestinal stromal tumours (GIST) - imatinib	Metastatic	Oct 2004
TA91	Ovarian cancer (advanced) - paclitaxel, pegylated liposomal doxorubicin hydrochloride and topotecan (review)	Advanced	May 2005
TA93	Colorectal cancer (advanced) - irinotecan, oxaliplatin and raltitrexed (review)	Advanced	Aug 2005
TA101	Prostate cancer (hormone-refractory) - docetaxel	Metastatic	Jun 2006
TA105	Colorectal cancer - laparoscopic surgery (review)	All stages	Aug 2006
TA110	Follicular lymphoma - rituximab	Advanced /	Sep 2006

TA Number	Title	Disease Stage	Date Issued
		Metastatic	
TA116	Breast cancer - gemcitabine	Metastatic	Jan 2007
TA118	Colorectal cancer (metastatic) - bevacizumab & cetuximab	Metastatic	Jan 2007
TA119	Leukaemia (lymphocytic) - fludarabine	All stages	Feb 2007
TA121	Glioma (newly diagnosed and high grade) - carmustine implants and temozolomide	Advanced	Jun 2007
TA124	Lung cancer (non-small-cell) - pemetrexed	Advanced / Metastatic	Aug 2007
TA129	Multiple myeloma - bortezomib	Advanced	Oct 2007
TA135	Mesothelioma - pemetrexed disodium	Advanced	Jan 2008
TA137	Lymphoma (follicular non-Hodgkin's) - rituximab	Advanced / Metastatic	Feb 2008
TA145	Head and neck cancer - cetuximab	Advanced	Jun 2008
TA162	Lung cancer (non-small-cell) – erlotinib	Advanced / Metastatic	Nov 2008
TA169	Renal cell carcinoma - sunitinib	Advanced / Metastatic	Mar 2009
TA171	Multiple myeloma - lenalidomide	Advanced	Jun 2009
TA172	Head and neck cancer (squamous cell carcinoma) - cetuximab	Advanced / Metastatic	Jun 2009
TA174	Leukaemia (chronic lymphocytic, first line) - rituximab	Advanced	Jul 2009
TA178	Renal cell carcinoma	Advanced / Metastatic	Aug 2009
TA176	Colorectal cancer (first line) - cetuximab	Metastatic	Aug 2009
TA179	Gastrointestinal stromal tumours - sunitinib	Advanced / Metastatic	Sep 2009
TA181	Lung cancer (non-small cell, first line treatment) - pemetrexed	Advanced / Metastatic	Sep 2009
TA183	Cervical cancer (recurrent) - topotecan	Metastatic	Oct 2009
TA184	Lung cancer (small-cell) - topotecan	Advanced	Nov 2009