NICE DSU TECHNICAL SUPPORT DOCUMENT 24: ADJUSTING SURVIVAL TIME ESTIMATES IN THE PRESENCE OF TREATMENT SWITCHING: AN UPDATE TO TSD 16

REPORT BY THE DECISION SUPPORT UNIT

produced in 2022/3, published in April 2024

Helen Bell Gorrod[1], Nicholas R Latimer[1], Keith R Abrams[2,3]

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

[2] Department of Statistics, University of Warwick, Coventry, UK

[3] Centre for Health Economics, University of York, York, 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

Website https://www.sheffield.ac.uk/nice-dsu

Twitter @NICE_DSU

ABOUT THE DECISION SUPPORT UNIT

The Decision Support Unit (DSU) is based at the University of Sheffield with members at the universities of York, Bristol, Leicester, Warwick and the London School of Hygiene and Tropical Medicine. 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 Centre for Health Technology Evaluation Programmes. Please see our website for further information: https://www.sheffield.ac.uk/nice-dsu

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.

ABOUT THE TECHNICAL SUPPORT DOCUMENT SERIES

NICE describes the methods it follows when carrying out health technology evaluations in its process and methods manual. This provides an overview of the key principles and methods of health technology assessment and appraisal for use in NICE Health Technology Evaluations. The manual does not provide detailed advice on how to implement and apply the methods it describes. The DSU series of Technical Support Documents (TSDs) is intended to complement the manual by providing detailed information on how to implement specific methods.

The TSDs provide a review of the current state of the art in selected topic areas. They make 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 evaluations, whether companies, 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. The TSDs will be amended and updated whenever appropriate. Where minor updates or corrections are required, the TSD will retain its numbering with a note to indicate the date and content change of the last update. More substantial updates will be contained in new TSDs that entirely replace existing TSDs.

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

Prof Allan Wailoo, Director of DSU and TSD series editor.

ACKNOWLEDGEMENTS

We would like to thank Donna Davis for support with formatting and proofreading. We also thank Rhiannon Owen and the reviewers from NICE (Summaya Mohammad and Lorna Dunning) for their extremely helpful contributions.

This report should be referenced as follows:

Bell Gorrod, H., Latimer, N. R., Abrams K.R. NICE DSU Technical Support Document 24: Adjusting survival time estimates in the presence of treatment switching: An update to TSD 16. 2024. [Available from nicedsu.org.uk]

EXECUTIVE SUMMARY

Treatment switching describes a situation where participants in a randomised controlled trial switch onto a treatment that they were not randomised to receive. Sometimes, treatment switches reflect treatment pathways that would be expected to be followed in standard clinical practice in the jurisdiction of interest, but in other cases they do not. Technical Support Document (TSD) 16 focused on a situation where patients randomised to the control group of a clinical trial are allowed to switch onto the treatment received by patients randomised to the experimental group.[1] Typically, this switching is not representative of a treatment pathway that would be expected to be followed in standard clinical practice, because the experimental treatment is usually new and not available at later lines of therapy. In such a situation, an intention to treat (ITT) analysis – whereby the data are analysed according to the arms to which patients were randomised – will not adequately address the decision problem faced in an appraisal of the new treatment: that of whether inserting the new treatment into the care pathway at its designated line of therapy represents a cost-effective use of resources. Instead, it may be necessary to estimate the effectiveness and cost-effectiveness that would be expected to be observed if treatment pathways used in standard clinical practice were followed – adjusting for the impact of treatment switches that would not be expected to occur in practice.

While it is often necessary to adjust when control group patients switch onto the experimental treatment, other types of treatment switching may also be unrepresentative of treatment pathways that would be expected to be observed in clinical practice, and further adjustment analyses may be required. For instance, patients in either randomised group may switch onto other experimental treatments, or onto treatments that are not available in standard clinical practice in the jurisdiction of interest, e.g. England and Wales.

TSD 16 described the treatment switching problem, and introduced a selection of adjustment methods that may be used – rank preserving structural failure time models (RPSFTM), iterative parameter estimation (IPE), marginal structural models (MSM) with inverse probability of censoring weights (IPCW) and two-stage estimation (TSE).[1-5] An analysis framework was presented, to help determine which adjustment

methods are likely to be appropriate on a case-by-case basis. Since TSD 16 was published a number of developments have occurred, motivating this new TSD.[1] These two TSDs should be read alongside one another and TSD 24 does not replace TSD 16 – instead it supplements and extends it.

Section 1 of this report provides an introduction and context. Section 2 describes the treatment switching problem, and includes a discussion around the different types of treatment switching and when adjustment may (or may not) be required. It is important to note that determining that it is relevant to adjust for a particular type of treatment switching does not mean that an adjustment analysis will be accepted for decision-making purposes. Treatment switching adjustment analyses can be prone to bias and error and while it may be relevant to adjust for a particular type of treatment switching, it may not be possible to do so reliably and robustly. Adjustment analyses should be assessed on a case-by-case basis, firstly with respect to the relevance of the analyses, and secondly with respect to the reliability and robustness of the analyses performed.

Section 3 summarises an addendum to the International Council for Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH) E9 Statistical Principles for Clinical Trials document that came into effect in 2020.[6] The ICH guidelines are heavily relied upon in the regulatory approval of new treatments, and the 2020 addendum to ICH E9 considers hypothetical estimands and refers to treatment switching, and is therefore highly relevant for Health Technology Assessment (HTA). We recommend that when planning trials and pre-specifying analyses, trial sponsors should consider questions of interest not only for regulatory agencies, but also for HTA agencies, and should pay particular attention to the implications that this can have for data collection.

In Section 4, we discuss approaches for addressing the problems associated with treatment switching both before and after a randomised controlled trial (RCT) has been run. We offer recommendations for trial planning, and provide a recap of the RPSFTM, IPCW and TSE adjustment methods,[2, 3, 5] and an overview of methodological developments that have occurred since publication of TSD 16. These include extensions to RPSFTM to allow for switching to treatments with different treatment effects,[7, 8] an increased capability to test the sensitivity of RPSFTM analyses to

violations of the common treatment effect assumption,[9] and a new version of the TSE method that uses g-estimation[10] and relies on fewer assumptions than the simple TSE method described in TSD 16.[1, 5] We also discuss issues relating to the practical application of IPCW, including covariate selection and extreme weights.[11]

In Section 5, we provide a review of recent NICE Technology Appraisals (TAs) affected by treatment switching, demonstrating that adjustment analyses are frequently reported in inadequate detail, and that treatment switching is often inadequately considered by external assessment groups and appraisal committees. This motivated the development of a set of recommendations and reporting guidelines, provided in Section 6, to improve the clarity and consistency of the reporting and review of treatment switching adjustment analyses in NICE TAs. Our reporting guidelines expand upon those recently proposed for the RPSFTM and IPCW methods by Sullivan *et al.* (2020).[12] The guidelines provide a list of summary statistics that should be provided for any trial that is affected by treatment switching, and a list of information to be presented alongside IPCW, RPSFTM, IPE, simple TSE, TSE with g-estimation, or other treatment switching adjustment analyses. We recommend that multiple treatment switching adjustment analyses should be performed to test the sensitivity of the results to the assumptions of the models used. We also recommend that external assessment groups and appraisal committees should more consistently consider the issue of treatment switching, both with respect to whether it is appropriate to adjust for the types of switching observed in pivotal RCTs, and with respect to the validity of any adjustment methods and analyses that are applied.

We also suggest areas for further research. In particular, it would be valuable to undertake neutral comparison studies that investigate more complex switching scenarios (for example, switching in both treatment arms, and to treatments with different treatment effects), which investigate the performance of all adjustment methods, including those developed more recently and extensions to previously existing methods. Further research on methods to adjust for treatment switching when only summary data are available would also be valuable.

CONTENTS

TABLES AND FIGURES

ABBREVIATIONS AND DEFINITIONS

AC	Appraisal Committee
ADT	Androgen Deprivation Therapy
AFT	Accelerated Failure Time
AG	Assessment Group
AIC	Akaike Information Criterion
ALK	Anaplastic Lymphoma Kinase
AZT	Azidothymidine
BIC	Bayesian Information Criterion
CDF	Cancer Drugs Fund
CI	Confidence Interval
DAG	Directed Acyclic Graph
dMMR	Deficient Mismatch Repair
ECOG	Eastern Cooperative Oncology Group
EGFR	Epidermal growth factor receptor
ERG	Evidence Review Group
FAD	Final Appraisal Determination
FO	Final Outcome
gBRCAm	Germline Breast Cancer Gene Mutated
HR	Hazard Ratio
HTA	Health Technology Assessment
ICH	International Council for Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use
IC3	Immune Cell 3
IPCW	Inverse Probability of censoring weights
IPE	Iterative Parameter Estimation
ISPOR	The Professional Society for Health Economics and Outcomes Research, formerly called the International Society for Pharmacoeconomics and Outcomes Research
ITC	Indirect Treatment Comparisons
ITT	Intention-to-treat
M2SM	Modified two-stage method
MSI-H	Microsatellite Instability-High

MSM	Marginal Structural Model
NHS	National Health Service
NICE	National Institute for Health and Care Excellence
NSCLC	Non-Small Cell Lung Cancer
OS	Overall Survival
PARP	Poly Adenosine Diphosphate-Ribose Polymerase
PD-1	Programmed Cell Death Protein 1
PD-L1	Programmed Cell Death Ligand 1
RCT	Randomised Controlled Trial
RPSFTM	Rank Preserving Structural Failure Time Model
SAP	Statistical Analysis Plan
SNFTM	Structural Nested Failure Time Models
TA	Technology Appraisal
TC3	Tumour Cell 3
TKI	Tyrosine Kinase Inhibitor
TSD	Technical Support Document
TSE	Two-stage Estimation

1. INTRODUCTION

Treatment switching is common in Randomised Controlled Trials (RCTs) used in submissions to the NICE technology appraisal (TA) programme. Switching occurs when patients in an RCT deviate from the treatment that they were randomly assigned to. It is often permitted due to ethical reasons and to aid trial recruitment, and is frequently seen in cancer trials, but it also occurs in trials in other disease areas.

This technical support document (TSD) provides an update to TSD 16 (Adjusting Survival Time Estimates in the Presence of Treatment Switching), published in 2014.[1] TSD 16 focused specifically on situations where patients randomised to the control group of an RCT switch onto the experimental treatment at some point during the trial. Subsequent papers have highlighted other types of switching in RCTs. In this TSD, we broaden our focus to include situations where participants randomised to the experimental arm of a trial switch onto the control treatment, or where participants randomised to either arm of the trial switch onto any other subsequent treatments. In Section 2, we provide further detail on this, and provide guidance and an over-arching recommendation on when adjustments for treatment switching may (and may not) be necessary. In particular, while in this TSD we broaden our focus to include switches onto any other subsequent treatments, this does not mean that it is appropriate to adjust for all these switches – usually it is only appropriate to adjust for switches onto non-standard treatments. In Section 3 we discuss the relevance for health technology assessment (HTA) and treatment switching of a recently published addendum to the ICH E9 Statistical Principles for Clinical Trials on Estimands and Sensitivity Analysis in Clinical Trials document, and provide a recommendation around the implications of this for trial planning in the context of HTA. In Section 4, we discuss methods for dealing with the treatment switching problem both before and after a trial has been run. First, we offer recommendations on considerations that should be made when planning a trial in which treatment switching is anticipated. Then we recap the treatment switching adjustment methods described in TSD 16, and summarise the methodological developments that have occurred since publication of the TSD.[1] In Section 5, we review recent NICE TAs affected by treatment switching, and in Section 6 we present a set of recommendations and reporting guidelines, with the aim of

improving clarity and consistency in the reporting (and review) of future treatment switching adjustment analyses.

As in TSD 16, our focus in this update document is on survival outcomes.[1] This is because treatment pathways become particularly important in economic evaluations that take a lifetime perspective – and in these, survival outcomes are often of critical importance. However, switching adjustment may also be relevant in a non-survival context, if trial outcomes are measured after treatment switches have occurred. For instance, treatment switching could affect subsequent measures of health-related quality of life, or other measures of clinical effectiveness. In addition, when adjustments to outcomes are made to account for treatment switches, adjustments to costs should also be included in economic analyses.

2. THE TREATMENT SWITCHING PROBLEM

What is treatment switching and when should we adjust for it?

RCTs aim to compare the effectiveness of a novel treatment (or combination of treatments) and a comparator. Patients enrolled in a trial are randomised to a treatment group. Treatment switching occurs when patients switch onto a treatment other than the one they were randomised to. An intention-to-treat (ITT) analysis represents the standard analysis used to establish the relative effectiveness of the treatment policies under investigation, by comparing trial outcomes in the randomised treatment groups. ITT analyses benefit from the important advantages associated with randomisation, and should always be presented.

The relevance of the ITT analysis in the context of HTA depends upon the decision problem faced, and sometimes other analyses may be required. If the treatment pathways followed in a clinical trial deviate from those that would be expected to be followed in clinical practice, trial outcomes measured on an ITT basis may not be useful for resource use allocation decision making. For example, imagine that patients randomised to the control group of a trial are allowed to switch onto the experimental treatment once their disease progresses, and that the experimental treatment has a beneficial impact on post-progression survival but is not available as a subsequent therapy in standard clinical practice. In that case, the patients in the control group may live longer than they would have if they had instead received a standard therapy. In such a case, the overall survival experienced in the control group of the clinical trial is lengthened by the switching, and an ITT analysis may over-estimate the effectiveness of the control treatment, and may result in an under-estimate of the cost-effectiveness of the experimental treatment. It would be desirable to adjust for the treatment switching in order to address the HTA decision problem and allow a fully-informed treatment recommendation to be made.

TSD 16 focused on switches from the control group to the experimental treatment.[1] However, we acknowledge that it is important to consider all types of treatment switching, and here we broaden the definition of “switching” to include switches between randomised treatments (from the control group to the experimental treatment

or from the experimental group to the control treatment), and switches from either randomised group to *any* other subsequent treatments.

While all treatment changes fall within the definition of "treatment switching", it is important to emphasise that not all switches need to be adjusted for in the context of a NICE TA. When determining which switches should be adjusted for, the decision problem being addressed must be considered. In NICE appraisals, the decision problem usually involves an assessment of whether inserting the new treatment into the care pathway at its designated line of therapy represents a cost-effective use of resources. Economic evaluations usually take a lifetime perspective when treatments impact survival, and therefore, whether modelled explicitly or implicitly, treatment pathways must be taken into account. If evidence on survival is taken from a clinical trial that is affected by treatment switching, the pathway of care received in the trial must be assessed, and it may be appropriate to move beyond ITT analyses to estimate the survival outcomes that would have been observed in the absence of the switching. Often it will be appropriate to adjust for some of the treatment switches that occurred in the trial, but not others. For example, it is likely to be appropriate to adjust for treatment switches that do not represent treatment pathways that would be expected to be followed in clinical practice, whereas it would not usually be appropriate to adjust for switches that do represent standard clinical pathways. An assessment of the treatments that it is appropriate to adjust for should be made on a case-by-case basis, but simple guidelines are provided below, followed by an over-arching recommendation:

- **Switches from the control group to the experimental treatment.** Adjustment for these switches is usually appropriate, because the experimental treatment is not usually available as a subsequent therapy as part of standard clinical practice in the jurisdiction for which a decision is being made. However, sometimes the treatment being appraised *is* already available as a later-line therapy, in which case it would not be appropriate to adjust for these switches.
- **Switches from the experimental group onto the control treatment.** Usually the control treatment represents a standard therapy which is available in standard clinical practice. In this case, it would not be appropriate to adjust for patients who switch from the experimental group onto the control treatment.

However, there may be situations where the control treatment is not available as a next-line therapy, in which case it may be appropriate to adjust for this type of switching.

- **Switches from either randomised group onto other treatments.** Here, the need for adjustment depends on whether the treatment switched to is available as part of the standard treatment pathway in the jurisdiction of interest. For NICE, if the treatment switched to is available at the relevant line of care in standard clinical practice in England and Wales, adjustment would not be required to address the HTA decision problem. In contrast, if a subsequent treatment is not available in standard practice at the relevant line of care, it would be appropriate to adjust. This should also be taken into account in the context of treatments available through managed access schemes, such as the Cancer Drugs Fund. It may be deemed appropriate to adjust for switches onto these treatments because whilst they may be available in the NHS, they are not part of routine commissioning.
- **Switch proportions.** Differing effectiveness of initial treatments may impact the proportion of patients that go on to receive subsequent treatments. In general, provided that treatment switches involve treatments that are representative of a standard treatment pathway, it is not necessary to adjust for different *proportions* of patients receiving subsequent treatments – unless it can be shown that the proportions are not compatible with what would be expected in standard clinical practice.
- **Timing of switches.** The timing of switches is important and may determine whether adjustments for switching need to be made. For treatments that are not part of the standard treatment pathway in the jurisdiction of interest, the timing of the switch is an important factor in analyses undertaken to adjust for the switching, but the timing does not alter the decision on whether or not it is appropriate to adjust for the switching. However, for treatments that do represent part of the standard treatment pathway in the jurisdiction of interest, the timing of the switch could impact whether adjustments are necessary. In particular, it is important to consider whether the switch times observed in the trial reflect the switch times that would be observed in standard clinical practice. For example, if switching in a trial is triggered by an interim analysis, rather than by a clinical event (such as disease progression, or toxicity), switching may

occur earlier (or later) in the trial than would be observed in reality. In this case, even if the treatment switched to represents part of the standard treatment pathway, it may be necessary to make adjustments to subsequently measured trial outcomes, due to the non-standard timing of the treatment switch. In this case, the analysis would seek to estimate what outcomes would have been observed if switching had taken place at the time at which it would be expected in standard clinical practice.

Recommendation 1: *It may or may not be appropriate to attempt to adjust for various different types of switching observed in an RCT in order to address the HTA decision problem. The switching observed in each treatment arm of a trial should always be described and it should be considered whether the observed switching is representative of treatment pathways available in standard clinical practice in the jurisdiction for which the decision is being made.*

It is important to note that determining that it is relevant to adjust for a particular type of treatment switching does not mean that an adjustment analysis will be accepted. As described in TSD 16, and in this document, treatment switching adjustment analyses are prone to bias and error.[1] It may be relevant to adjust for treatment switching, but it may not be possible to do so reliably and robustly. Adjustment analyses should be assessed on a case-by-case basis, firstly with respect to the relevance of the analyses, and secondly with respect to the reliability and robustness of the analyses performed.

In addition, it is only necessary to adjust for switching if switching is expected to impact upon the outcomes being measured. For example, if patients in a trial switched onto a palliative treatment that is not available in clinical practice, it unlikely to be necessary to adjust survival outcomes. It may, however, be relevant to adjust cost and quality of life estimates.

3. TREATMENT SWITCHING AND THE ICH E9 (R1) ADDENDUM

The International Council for Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH) bring regulatory bodies and the pharmaceutical industry together to develop guidelines relating to quality, safety and efficacy to ensure that high standards are met whilst efficiently assessing medicines. These are relevant for HTA because the analyses used in submissions to HTA agencies are often the same as those presented to regulatory bodies. Therefore, any changes made to ICH guidelines may have important implications for analyses submitted to NICE.

The ICH E9 Statistical Principles for Clinical Trials was initially published in 1998, with the aim of harmonising statistical methods used to analyse trials. In 2020, an addendum to ICH E9, on *estimands* and sensitivity analysis, came into effect.[6] The addendum highlights the complexities associated with estimating treatment effects in the presence of intercurrent events, which are events that occur post-randomisation and prior to the end-point of interest and which have implications for the estimation and interpretation of treatment effects. Treatment switches are described as examples of intercurrent events, representing the first time that ICH guidelines have referred specifically to analyses that may adjust for treatment switches. Given the prominence of ICH guidelines, and the importance of treatment switching for HTA, it is important for those involved in HTA to understand the ICH E9 addendum. Here, we provide a brief overview of the addendum with respect to trial planning and estimation strategies.

3.1 ICH E9 ADDENDUM: TRIAL PLANNING

The addendum advocates carefully defining an estimand to address the research question of interest at the planning stage of the trial, by considering the following five aspects;[6]

- **Treatment**, including identification of the study treatment, comparator and potential subsequent treatments.

- **Population**, that is the relevant patient group for the clinical question of interest, inclusion/exclusion criteria of the trial, and expectations for the real-world population.
- The **variables or endpoint** required to answer the clinical question of interest. The addendum states that it is important to collect all data that are needed to support estimation of the estimand(s) of interest. Trials should be planned to allow this. In this TSD, we focus on overall survival as the end point of interest, and note that data collection on patient characteristics and treatment switches are required to apply statistical adjustment methods to adjust for treatment switching (see Section 4). As previously acknowledged, other outcome measures may also be affected by treatment switching, and the data collection required will depend upon the clinical questions of interest / endpoints identified at the trial planning stage.
- List of potential **intercurrent events** and plans for addressing each type of intercurrent event. The addendum notes that treatment switches can represent intercurrent events, and five estimation strategies are suggested for addressing these events (see 3.2, below).
- The **population-level summary statistic** for the comparison of treatments. It is relevant to note that hazard ratios (HR) are usually used to compare survival outcomes in clinical trials, but, for the purposes of HTA, entire survival curves are of most importance. It is important to note that most statistical methods that may be used to adjust for treatment switching provide adjusted survival times ("counterfactuals") or weighted survival times, and these can be used in any survival model (see Section 4). Therefore, modellers are not restricted to using estimated HRs when incorporating adjustment for treatment switching in economic evaluations. While extrapolation is almost always required when assessing the cost-effectiveness of interventions that affect survival, it is relevant to note that restricted mean survival time (RMST) can be used to summarise survival benefits observed during the period of a clinical trial and this might be more meaningful than a hazard ratio, particularly in the presence of non-proportional hazards.[13] RMST adjusted for treatment switching can be estimated using the adjusted or weighted survival times produced by the adjustment methods described in this TSD.

3.2 ICH E9 ADDENDUM: ESTIMATION STRATEGIES

The ICH E9 addendum suggests five estimation strategies that can be considered to address intercurrent events.[6] Of these strategies, the **Treatment Policy** and **Hypothetical** strategies are the most relevant in the context of treatment switching.

- The **Treatment Policy** strategy describes a comparison of treatment groups as randomised, i.e. an ITT analysis. In the context of NICE appraisals, this type of analysis will be considered appropriate if all switches are made to approved subsequent therapies, or there are very few switchers (causing the effect of switching to be negligible).
- The **Hypothetical** strategy describes an estimation of what would have happened if the specific intercurrent events (such as specific treatment switches) had not happened. The treatment switching adjustment methods outlined in Section 4 can be used to obtain estimates of survival in the absence of treatment switching.

For survival outcomes in trials affected by treatment switching in NICE appraisals, we recommend that results are presented under both the treatment policy (ITT) and hypothetical (treatment switching adjusted) strategies, as described in the reporting guidelines in Section 6.

The three other estimation strategies specified in the addendum are referred to as “Principal stratum”, “While on treatment” and “Composite” strategies. These are less relevant in the treatment switching context. The principal stratum strategy requires the analysis of sub-populations within the trial, defined by whether or not the intercurrent event would occur. In HTA, the objective is to estimate the effectiveness and cost-effectiveness of the experimental treatment in the entire eligible population. As described in the NICE manual,[14] it is often relevant to consider subgroups, but it is unlikely to be appropriate for these subgroups to be defined according to whether or not a patient would be expected to switch treatment during follow-up of a trial. This is because treatment switching is unlikely to be random, with switching often motivated by disease progression or lack of response – therefore, analysing subgroups of patients based on whether they would be expected to switch or not involves conditioning on a future event and is likely to introduce bias. In some cases, a principal

stratum strategy might be informative – for example, an estimate of the treatment effect in patients able to tolerate a treatment might be useful. However, in general, this strategy seems less likely to be relevant in an HTA context, including when adjustments for treatment switching are required.

The while-on-treatment strategy only considers events that occur prior to the intercurrent event. It may be useful when, for example, the outcome of interest is adverse drug reactions, but some participants discontinue treatment: it may then be appropriate to assess the event risk while participants remain on treatment. However, it is not appropriate in the context of treatment switching and the estimation of treatment effects on survival, because the strategy would involve censoring data for patients who switch at the time of switch, which is highly prone to selection bias as switching is likely to be correlated with prognosis.

The composite strategy requires that information on the intercurrent event is incorporated into the endpoint variable, thus changing the variable from a binary endpoint to a categorical variable. This might be helpful in some contexts – for example, if the primary outcome in a trial was response or non-response, but some patients discontinued treatment before response was measured, then the intercurrent event of treatment discontinuation could be added to the outcome variable. Progression-free survival is itself a composite endpoint variable, measuring a combination of the growth of a tumour and survival. However, in the context of treatment switching and OS, a variable that combines treatment switch and survival events would not be intuitively meaningful, and would not be appropriate for addressing HTA decision problems – particularly where economic models are used, splitting the disease pathway into mutually exclusive health states.

See the addendum itself,[6] Manitz *et al.* (2022) and Clark *et al.* (2022) for further discussion of the estimation strategies and examples of estimands in the context of treatment switching.[15, 16]

3.3 ICH E9 ADDENDUM: IMPLICATIONS FOR HTA

By acknowledging treatment switching as an intercurrent event, the ICH E9 addendum may result in adjustment analyses becoming more commonplace in regulatory submissions – in the past these analyses were frequently conducted on an *ad hoc* basis specifically for HTA submissions. The addendum provides a basis for estimating treatment effects moving beyond the ITT principle, but is clear that appropriate estimands and analyses should be pre-specified at the trial planning stage, and that trials should be designed in a way that allows reliable estimates to be obtained for relevant estimands, for example by incorporating appropriate data collection strategies. The addendum does not describe statistical adjustment methods in detail, but acknowledges that some estimands may require methods of analysis that involve important assumptions, and trials should be designed with these in mind. As described in Section 4 of this TSD, several adjustment methods require intensive data collection, and the ICH E9 addendum specifically states that all relevant data should be collected to support estimation. The addendum also states that results from all analyses should be reported systematically, specifying whether each analysis was pre-specified.[6] Through improved pre-specification of treatment switching adjustment analyses, improved trial design to support the subsequent application of adjustment methods, and transparent reporting of analyses conducted, adoption of the ICH E9 addendum may improve the quality and robustness of adjustment analyses included in submissions to HTA agencies.

Recommendation 2: *We recommend that when planning trials and pre-specifying analyses, trial sponsors should consider questions of interest not only for regulatory agencies, but also for HTA agencies, and especially the implications that this can have for data collection. Recommendations made in the ICH E9 addendum around trial planning, pre-specification of analyses, and reporting of results, should be adopted.*

4. TREATMENT SWITCHING ADJUSTMENT: PLANNING, METHODS, AND DEVELOPMENTS

In this section we give an overview of how to deal with the treatment switching problem both before and after a trial has been run. Our guidance on trial planning (in Section 4.1) is based upon our knowledge of the requirements of adjustment methods and difficulties that we have observed in the application of adjustment methods when treatment switching has not been anticipated or planned for. Our overview of adjustment methods (Section 4.2) includes a re-cap on the main characteristics of the methods, together with a summary of methodological developments that have occurred since the publication of TSD 16.[1] This summary is based on a targeted review of the literature – methods for this review are presented in Appendix A.

4.1 PLANNING A TRIAL

Treatment switching should ideally be considered at the planning stage of every trial. Here, we build on the advice provided by Henshall, *et al.* (2016).[17] The recommendations listed below should be taken into account alongside the development of an estimand, as referred to by the ICH E9 (R1) Addendum.[6] Here we focus on aspects specifically relevant to treatment switching – we do not cover some more general issues, such as internal validity of the data collected (including data curation, quality insurance), which should be taken into account when analysing any dataset.

Recommendation 3: *Consider the types of switches that could occur*
Switches from the control arm to the experimental treatment should be considered, as well as switches from either arm to subsequent treatments which do not represent standard care. It is important to acknowledge that what constitutes a “non-standard” subsequent treatment will differ across countries and jurisdictions, and so different adjustment analyses may be required in different jurisdictions.

Recommendation 4: *Consider when the switches could occur*
Some trials will specify in the protocol that treatment switching will be permitted only after disease progression, or some other clinical event, or on the basis of interim

results. The time of earliest switch will affect which methods can be applied, for instance the simple version of the two-stage estimation (TSE) adjustment method can only be applied if switching occurs *after* a disease related secondary baseline.

Recommendation 5: *Consider sample sizes*

The trial should recruit sufficient participants to allow conclusions to be drawn on key outcomes and end points even if treatment switching occurs. It should be noted that treatment switching adjustment methods are less likely to perform well when sample sizes are small and switching proportions are high. Potentially, sample size calculations could formally include assumptions around treatment switching, though this represents an area where further research would be valuable.

Recommendation 6: *Consider data collection*

Different treatment switching adjustment methods have different data requirements. For instance, the rank preserving structural failure time model (RPSFTM) and iterative parameter estimation (IPE) methods require data on assigned treatment arm, time to death, occurrence of switch and time of switch, whereas the inverse probability of censoring weights (IPCW) and TSE methods require additional data on patient characteristics at regular time intervals to meet the “no unmeasured confounders” assumption.[1-4, 10, 18] This assumption ideally requires that all characteristics that both have a direct effect on the decision to switch *and* the probability of death are collected. Although the no unmeasured confounders assumption may seem impossible to meet perfectly, it should be kept in mind that the decision to switch can only be based on the information that the clinician has available, combined with the clinicians own clinical experience. Missing data (or unmeasured variables) are not necessarily a problem, unless the information is observable by the clinician and/or patient, but was not collected in trial datasets. It is useful to set out the causal relationships between relevant variables in the form of a Directed Acyclic Graph (DAG), as a tool for informing appropriate variable selection and data collection.[19, 20] This is discussed in more detail in subsection 4.2.3.2.

Recommendation 7: *Describe the plans for adjusting for treatment switching in the statistical analysis plan (SAP)*

Plans for the application of treatment switching adjustment methods should be described in the Statistical Analysis Plan (SAP) for the trial, and this should be set out prior to the trial initiation. The reporting guidelines in Section 6 of this TSD stipulate that appropriate adjustment methods should be applied and compared. Each method makes strong assumptions, so ideally a range of adjustment methods should be applied that rely on different assumptions. It is unlikely to be possible to pre-specify the most appropriate adjustment method, but it is possible to specify analyses that will be undertaken to provide information that will help determine which method has produced the most reliable results – for example, by assessing the range of IPCW weights and the performance of g-estimation. These analyses are detailed in the reporting guidelines in Section 6. Once the trial data have been collected, the statistical methods should be applied to the data as described in the SAP.

4.2 ADJUSTMENT METHODS

In this section, we provide a brief re-cap of the methods described in TSD 16, followed by a summary of new developments. Simple methods, such as censoring at the time of switch or excluding switchers from the analysis, are prone to bias and are not recommended. Switching is often related to prognosis, therefore excluding or censoring switchers removes a relevant subset of the trial population and creates a non-random loss of information. To appropriately address the NICE TA decision problem, more complex methods are required. We therefore focus on more complex methods that adjust for treatment switching. The adjustment methods presented in TSD 16, including IPCW, RPSFTM, IPE, and TSE, remain relevant.[1-3, 5]

4.2.1 RPSFTM and IPE

The standard RPSFTM can be applied in situations where patients switch between the control and experimental treatments, but is not applicable to situations that involve a switch to a subsequent treatment which does not represent standard care.[2] The application of the model requires dividing observed patient time into time spent on the control treatment (T_i^c) and time spent on the experimental treatment (T_i^e). For those that do not switch to the experimental treatment, T_i^e is zero. A common treatment effect is assumed for all patients regardless of when the treatment is received – that is, the time ratio (ψ) (sometimes referred to as the “acceleration factor”) associated with

receiving treatment is the same for all patients (relative to the amount of time the treatment was taken for), whether they were originally randomised to the experimental treatment or only switched onto it later. Counterfactual survival times U_i are specified as follows

$$U_i = T_i^c + exp(\psi)T_i^e \qquad (1)$$

The RPSFTM method identifies an optimal value of ψ using g-estimation, which involves entering a range of values for ψ into the counterfactual model and using a g-test to compare the counterfactual survival in randomised groups.[2] The choice of test used for the g-test can affect results and should be justified – typically a log-rank test is used but other options are available, such as a Wilcoxon test, which gives more weight to earlier time-points. The optimal value of ψ is found when the g-test z-statistic equals zero, i.e. where counterfactual (untreated) survival times in the control group are equal, on average, to counterfactual (untreated) survival times in the experimental group. In some instances, there may be multiple values for ψ that result in a zero z-test statistic. In such circumstances some software packages include an algorithm so that an average of the solutions for ψ is calculated,[9] but it is always important for the analyst to investigate the data and the g-estimation output to identify whether multiple solutions exist, and to determine why this might be the case. Alternatively, the IPE method could be used to estimate ψ.[4] The IPE method is based upon the same counterfactual framework as the RPSFTM and relies on the common treatment effect assumption, but instead of using g-estimation to identify an optimal value of ψ, the IPE method iteratively fits a parametric survival model. This requires an additional assumption that is not required by the RPSFTM – that is, that survival times follow a parametric distribution. However, taking the IPE approach has the advantage of ensuring that multiple solutions will not be identified for ψ.

Once the value of ψ is obtained – whether using the RPSFTM or the IPE method – survival times in switchers can be adjusted to provide the hypothetical survival times that would have occurred if there had been no switching. A new dataset is then obtained, and survival analysis can be undertaken on it – this could include fitting standard parametric models to the adjusted data, or more flexible models if standard models are deemed inappropriate (see TSD 21).[21] However, when fitting models to

the adjusted datasets provided by RPSFTM or IPE adjustment, the fact that the data are no longer fully observed (that is, the data are adjusted) must be taken into account. To adequately capture the uncertainty associated with this, the entire adjustment and subsequent model-fitting process should be bootstrapped, or, if deriving an adjusted hazard ratio, by using the p-value from the ITT analysis.[2, 22]

If censoring is present in the data, re-censoring can be applied to adjust the data in all groups affected by treatment switching.[23] Re-censoring aims to break the dependence between treatment received, counterfactual censoring time and prognosis, which may cause bias in the adjusted dataset – see Section 4.2.1.1 for more details. The recommendations in Section 6 of this TSD state that the RPSFTM and IPE (and TSE) methods should be applied with and without re-censoring for submissions to NICE.

The RPSFTM can be applied using the strbee package in Stata[9] or the RPSFTM package in R.[24] The IPE can also be applied using the strbee package in Stata.[9] It is notable that both methods can be applied on an "on treatment" (sometimes referred to as "as treated") or a "treatment group" (sometimes referred to as "ever treated") basis. The "on treatment" approach only considers time actually spent receiving the experimental treatment as time spent on treatment (i.e. time during which the T_i^e indicator equals 1). The "treatment group" approach considers all time after initiating the experimental treatment as time spent on treatment (i.e. the T_i^e indicator equals 1 for all time periods after treatment initiation). The "treatment group" approach may be useful when a treatment effect beyond treatment discontinuation is expected, or when the control treatment is an active therapy (in which case the estimate of ψ will represent the effect of the treatment pathway initiated when starting the experimental treatment, rather than the effect specific to the experimental treatment itself). See Section 3.3.1 of TSD 16 for more information on this.[1]

4.2.1.1 Extensions and recent developments to RPSFTM and IPE

Issues relating to re-censoring

Censoring is problematic in the counterfactual datasets created when adjusting for treatment switching, because whilst adjusted event times are estimated for people who switched treatments and experienced an event (usually death), adjusted censoring

times are estimated for people who switched treatments and did not experience the event of interest during trial follow-up. This introduces a relationship between switching and censoring times, because censoring times are adjusted in switchers but not in non-switchers. Because switching is likely to be associated with prognostic characteristics, a relationship is created between prognostic characteristics and censoring times, resulting in informative censoring. The purpose of re-censoring is to break the dependence between switching, censoring time and prognosis, and involves estimating adjusted potential censoring times for *all* patients in a treatment arm in which switching occurred, not just for patients who switched.

Re-censoring results in valid estimates of treatment effects up to the maximum recensored follow-up time. However, re-censoring involves a loss of longer-term data, and the larger the treatment effect, the more data are lost, because re-censoring times are a function of the estimated treatment effect. When the intention is to extrapolate survival beyond the period of the trial, the loss of information associated with re-censoring could result in poor extrapolations and biased estimates of long-term treatment effects. This is likely to be the case when important changes in hazards occur beyond the recensored follow-up times, and/or when important changes in the treatment effect occur beyond the recensored follow-up times. Latimer *et al.* (2019) performed simulations to study the impact of re-censoring and found that analyses both with and without re-censoring were prone to bias, and concluded that including re-censoring should not always represent the default approach when estimating long-term survival. This is of particular importance, given that a key objective in HTA often involves estimating life-time survival benefits, and these estimates are frequently a key driver of cost-effectiveness results. The authors found that the two approaches often provided bias in opposite directions, and therefore recommended that analyses should be performed with and without re-censoring, as this may give decision-makers a clearer idea of the range in which the true longer-term treatment effect lies.[25]

The original paper that introduced the IPE method, published by Branson and Whitehead (2002),[4] described an abridged version of re-censoring, which was subsequently shown to be sub-optimal.[26] Since publication of TSD 16, Zhang and Chen (2016) have further discussed re-censoring in the context of the IPE, and

proposed a “modified IPE” which applies re-censoring correctly.[27] This is in line with recommendations previously made in TSD 16.

Weighted log-rank test for RPSFTM

The log-rank test that is used within the RPSFTM performs optimally when the hazard ratio comparing control and experimental group survival times is constant over time. However, in the presence of an experimental treatment with a non-zero treatment effect, treatment switching results in changes in the ITT-estimated hazard ratio over time, even if the actual treatment effect is constant over time. To address this issue, Bowden *et al.* (2016) propose a weighted log-rank test, where weights are derived from the proportions of trial participants on the experimental treatment in each trial arm over time.[28] The authors propose the use of these weights for testing the ITT null hypothesis of no difference in the survival distributions between randomised groups when treatment switching is present, and further propose their use within the RPSFTM framework. The authors demonstrate that a simple weighted log-rank test can be statistically more powerful than a standard log-rank test when substantial treatment switching occurs, and when large rates of separation between the two randomised treatment groups’ survival functions over time coincide with large differences in the proportion of patients “on treatment” in each arm. Thus, using a weighted log-rank test could result in narrower confidence intervals around estimated treatment effects in ITT and RPSFTM-adjusted analyses in the presence of treatment switching. Bowden *et al.* (2016) suggest that the RPSFTM with a weighted log-rank test should be used alongside the standard RPSFTM, as a supplementary analysis.[28] However, the authors also caution that using a weighted log-rank test could result in inappropriate conclusions around treatment effectiveness if in reality an experimental treatment provides early benefit and later harm (or reduced benefit), because more weight would be given to earlier event times. R code is available from the authors. A modified log-rank test (Jimenez *et al*. 2021) and a weighted log-rank method (Ristl *et al.* 2020) were also identified by our literature searches as amended versions of the log-rank test that may also be considered as supplementary analyses in the presence of treatment switching.[29, 30]

Adjusting for switches to more than one treatment, or to a treatment with a different effect

The common treatment effect assumption is crucial to the standard one-parameter RPSFTM. The model assumes that patients are either "on" or "off" treatment, and therefore only one treatment effect is estimated. Hence, the method is not appropriate when the objective is to adjust for switches to "other" treatments – i.e. when switching is to a treatment other than one of the randomised treatments, or, more specifically, when switching is to a treatment that has a different effect to those associated with the randomised treatments. Over time, various studies have attempted to extend the RPSFTM method by re-parameterising the model to allow more than one treatment effect to be estimated.[23, 31] However, these attempts have been unsuccessful, with authors finding that bivariate RPSFTM models do not provide robust results.

Xu *et al.* (2021), propose a stratified RPSFTM which allows for patients in the control arm to switch to multiple treatments that may or may not be part of the original trial protocol.[7] The authors refer to this as "multilevel switching", which could include different doses of one of the randomised treatments, or completely different treatments. Their model reflects the RPSFTM as described in equation (1), with the difference being that a series of parameters ψ_k, $k=0,\ldots,K$ are estimated, which represent the different "levels" of treatment, instead of a single parameter for ψ. An advantage of the model is that it relaxes the common treatment effect assumption to allow for different effects for each treatment level. However, the model is limited by the extra complexity that is introduced by incorporating additional parameters. This is associated with increased computational burden, potential difficulties in obtaining a unique solution, and possible biases in parameter estimates. The authors use the same g-estimation procedure for multiple acceleration parameters described (and shown to be imprecise) by Robins and Greenland (1994).[31] The difference between the Xu *et al.*'s (2021) stratified RPSFTM and the method proposed by Robins and Greenland (1994) is that Xu *et al.* adjust for a switch from the control group to different levels of the experimental treatment and explicitly use RPSFTM, whereas Robins and Greenland (1994) adjust for a switch to an alternative treatment from randomly assigned low or high doses of the experimental treatment, and use a more general structural nested failure time model. Despite these superficial differences, the models appear to be very similar and it is unclear whether the method of Xu *et al.* (2021)

provides any advantages over the Robins and Greenland (1994) method. Given the similarities between the methods, we would expect them to be subject to the same limitations – i.e. in practice it is not possible to reliably estimate multiple parameters using the RPSFTM method without making additional assumptions.

Li *et al.* (2017) present what they describe as an "enhanced" RPSFTM, to adjust survival estimates for post-disease progression treatment switching from the control group to the experimental treatment, as well as for switches to an alternative subsequent treatment, by introducing a second parameter into the model and making additional assumptions.[8]

$$U_i(\psi_1, \psi_2) = T_i^{PFS} exp(Z_i^0 \psi_1) + \left(T_i - T_i^{PFS}\right) exp(I(Z_i^P = 2)\psi_2 + I(Z_i^P = 1)\psi_1) \qquad (2)$$

U_i represents the counterfactual survival time for patient *i*, ψ_1 and ψ_2 represent the treatment effect 1 and treatment effect 2, T_i denotes overall survival time, T_i^{PFS}denotes progression-free survival time, Z_i^0 represents treatment received at randomisation and Z_i^Prepresents treatment received post-progression. The authors apply additional assumptions beyond those specified by Robins and Greenland (1994) to allow two treatment effects to be estimated. Li ***et al's*** model is designed specifically for a case where switching is either to the experimental treatment, or to one other subsequent treatment. Whilst the model refers to progression-free survival, it is actually more general, in that treatment effects for the second-line treatments are estimated for the post-switch period – that is, the actual time of disease progression is not used in the model; instead the time of switch is assumed to equal the time of disease progression. This is important because the authors go on to estimate the post-progression treatment effect associated with the treatments that may be switched to and, if actual switching times varied with respect to disease progression times, these second-line treatment effects may become biased (as for the simple TSE method – see Section 4.2.3).

The premise of Li ***et al.'s*** approach is that the standard RPSFTM can only reliably estimate one treatment effect parameter (in line with the original findings of Robins and Greenland (1994) [31]), and so it is necessary to use a two-stage procedure when

attempting to estimate two treatment effects. Their proposed method achieves this by assuming that the treatment effects ψ_1 and ψ_2, are related through a nuisance parameter, γ, where $\gamma \equiv \psi_2 - \psi_1$. Thus, if γ can be estimated, it can be substituted into $\psi_1 + \gamma$ to represent ψ_2, allowing g-estimation to be used as in the standard RPSFTM to estimate ψ_1. To estimate γ, the authors assume that post-progression survival times are distributed according to a Weibull distribution with a constant shape parameter, κ, and a baseline (no therapy) scale parameter, λ, which is a function of ψ_1 and ψ_2, and depends on post-progression treatment Z_i^P and patient characteristics at baseline and disease progression. By fitting a Weibull model to post-progression survival times, controlling for baseline characteristics and those measured at disease progression (i.e. the time of switch), comparing those who switched onto the experimental treatment and those who switched onto the alternative treatment, γ can be estimated. The authors further propose combining their enhanced RPSFTM with IPCW to adjust for any bias arising from informative drop-out that may occur during the trial.

Li *et al.* present a simulation study comparing their enhanced RPSFTM method against ITT, per protocol, censoring at time of switch and a Cox regression with time-varying treatment covariates. The authors demonstrate that the enhanced RPSFTM method produces less bias in the estimated treatment effect compared to these other methods.[8] The simulation study indicates that the solutions for γ in the proposed model were unstable when there were a small number of patients who remained in the study after disease progression. The authors recommend that a value of γ based on the literature and expert opinion could be used in these circumstances. In addition, they recommend keeping the number of treatments, or levels of treatment, to a minimum, due to the increased computational burden arising with each additional parameter added to the model.

Li *et al.'s* method is innovative, but its performance against other methods that can be used to adjust for switches to various subsequent treatments (e.g. IPCW and TSE) is unclear. The two stages involved in the method mean that it has similarities to the TSE method, but it is designed specifically for cases where two types of switching occur – from the control group onto the experimental treatment; and from the control group

onto an alternative treatment – with the emphasis of the first stage of the method being on estimating the difference in effectiveness between these two types of switching. As discussed in Section 4.2.3, the TSE method can also be used to adjust for switches to more than one treatment. Like the TSE method, Li ***et al.'s*** method requires information on patient characteristics at baseline and at the time of progression (or switch), and requires the assumption of no unmeasured confounding for the estimation of γ. The authors do not address issues around time-dependent confounding that could occur if switches do not all occur exactly at the time of disease progression. In addition, owing to their use of the RPSFTM and g-estimation to estimate the effect of the experimental treatment, Li *et al.* assume that the effect of the experimental treatment is the same before and after disease progression (i.e. the common treatment effect assumption) in order to keep the dimension of the parameter space to 2, rather than 3 – an assumption that the TSE approach does not rely on. Therefore, Li *et al.*'s method requires *both* the no unmeasured confounding *and* the common treatment effect assumptions to hold.

Relaxing the common treatment effect assumption

In response to concerns about the plausibility of the common treatment effect assumption, software packages that implement RPSFTM have been modified to allow the treatment effect to vary between individuals. The strbee Stata package includes a 'psimult(k)' option, and the rpsftm R package includes a 'treat_modifier(k)' option.[9] The default is for k=1, but if, for example, k was set to 1 in patients randomised to the experimental arm of the trial, and was set to 0.7 in the control arm, the RPSFTM analysis would be run under the assumption that the effect of the experimental treatment is 30% lower in patients from the control group who switch onto the experimental treatment, than in patients initially randomised to the experimental group. In this way, k could be varied between, for example, 0.5 and 1.5, to show how sensitive the RPSFTM analysis is to the common treatment effect assumption, allowing for the potential that the experimental treatment could be 50% less effective (or 50% more effective) in switchers.[9]

This technique for relaxing the common treatment effect assumption differs importantly from the multi-parameter versions of RPSFTM previously investigated by Robins and Greenland (1994) White *et al.* (1999), and Xu *et al.* (2021).[7, 23, 31] The 'psimult' and

'treat_modifier' options still only estimate one value of ψ, but assume that the actual treatment effect received by defined groups (such as those initially randomised to the experimental treatment and those in the control group who switch onto the experimental treatment) is a multiplicative factor of ψ. This avoids problems associated with having to estimate more than one treatment effect. The limitation is that the analyst has to specify the multiplicative factor(s), making it/them somewhat arbitrary. However, by running scenario analyses with a series of multiplicative factors (e.g. 0.5, 0.6, 0.7, 0.8, 0.9, 1.0 and perhaps greater than 1.0 if it is considered possible that switchers could benefit more from the experimental treatment than those initially randomised to the experimental group), useful information on the sensitivity of the RPSFTM analyses to the common treatment effect assumption can be obtained. Deciding on suitable values for these multiplicative factors represents an area where the elicitation of clinical expert opinion would be valuable. Given that the common treatment effect assumption is sometimes perceived to suffer from a lack of plausibility, this development in the application of the RPSFTM is valuable. This sensitivity analysis allows the impact of violations of the common treatment effect assumption to be quantified, allowing interpretation of the results of RPSFTM analyses to be better informed. This may allow decision-makers to use RPSFTM analyses to inform their recommendations more confidently, as opposed to a situation where the impact of violations of the common treatment effect assumption is completely unknown.

However, it remains likely that the RPSFTM will only be appropriate when switching is directly between randomised groups, as assigning the effect of subsequent treatments to be a multiplicative factor of the experimental group treatment effect would appear to make most sense when the treatment switched to is in fact the experimental treatment (or perhaps a drug of the same class). Alternatively, each subsequent treatment could be assigned a value of *k* such that its effect is a multiplicative factor of the experimental treatment effect. This would be similar to the approach suggested by Li *et al.* when the objective is to estimate multiple treatment effect parameters – that is, using the literature and expert opinion to estimate the effect of each subsequent treatment relative to the experimental treatment.[8] However, this would involve making very strong assumptions about the relative effects of the various subsequent treatments and moves far beyond analysing the data from the pivotal trial to estimate treatment effects. Such an approach may be more credible in circumstances where

evidence on the relative effectiveness of the relevant subsequent treatments is available.

4.2.2 IPCW

Inverse probability of censoring weights (IPCW) can be applied in marginal structural models (MSM) to adjust for treatment switching from the control group to the experimental treatment, and from either trial arm to other subsequent treatments.[3, 20] The approach requires a "switching model" to estimate the weights, and an "outcomes model" to estimate the treatment effect adjusted for switching. The outcomes model (also referred to as an MSM) could take the form of a Cox model, to estimate a hazard ratio, or parametric models could be fitted to IPCW weighted survival times to perform extrapolation. The method relies upon the no unmeasured confounding assumption, and therefore requires data for each patient on prognostic characteristics that influence the probability of switch *and* survival (or another outcome of interest). A positivity assumption is also required, which specifies that there are no confounding factors that perfectly predict switching.[3, 11]

To apply IPCW, data on prognostic covariates are required at baseline and at regular time intervals until the earliest time-point of either switch, death, drop-out or end-of-follow-up. For patients who switch, the data are artificially censored at the time of switch. This can create bias due to informative censoring, because switching is likely to be associated with prognostic characteristics. To address this, a weight for each remaining patient at each time point is calculated based upon estimated switch probabilities. Patients who have similar prognostic characteristics to switchers, but who themselves have not yet switched, receive a weight that is greater than 1. Provided the model used to estimate switch probabilities is appropriately specified, and the no unmeasured confounding assumption holds, this weighting corrects for the bias introduced by censoring switchers. Survival analysis can then be undertaken on the weighted dataset and, as for the RPSFTM, standard or more flexible survival modelling can be used (see TSD 21 [21]).

The switching model estimates the probability of not switching for each individual over time, conditional on prognostic baseline and time-dependent variables. Time-

dependent weights are estimated for each individual, with the weight representing the inverse probability of remaining unswitched over time – that is, the probabilities estimated by the switching model represent the denominator of the weight. These are "unstabilised" weights and can be highly variable. For this reason, "stabilised" weights are often used. The switching model for stabilised weights involves the estimation of a model for the numerator of the weight in addition to the model used for the denominator for unstabilised weights. The model use for the numerator of stabilised weights can be specified in different ways, but must not include any time-dependent confounding variables. Typically, the model used for the numerator is the same as that used for the denominator, but includes only prognostic baseline variables.[10] Any prognostic baseline variables that are included in the numerator of the weighting model should also be included in the weighted outcomes model. While stabilised weights are likely to result in less extreme weights, unstabilised weights are more intuitive to apply to survival models for extrapolation, and more straightforward to interpret when estimating treatment effects (with respect to conditional and marginal effects).

4.2.2.1 Extensions and recent developments to IPCW

Our review did not identify specific extensions to the IPCW method, but research relevant to the method has been published. In our experience, key issues around the application of IPCW surround covariate selection and extreme weights, so we focus mainly on these issues here.

Adjusting for multiple types of switching

If treatment switching to subsequent treatments occurs in both arms of a randomised trial, IPCW can be applied to both arms. Separate weighting models should be estimated for each arm of the trial, and used to generate a weight per patient at each time point. Separate weighting models could also be specified for each treatment that patients switch to within each randomised arm, but it is unclear whether this is preferable to simply grouping together all treatments that adjustment is required for in each treatment arm – further research on this would be valuable.

Variable selection and directed acyclic graphs (DAGs)

Selecting appropriate variables for inclusion in the IPCW switching models (and the outcomes model) is crucial in obtaining unbiased adjusted outcomes. The positivity assumption and no unmeasured confounders assumption are intrinsically related to data availability and variable selection.

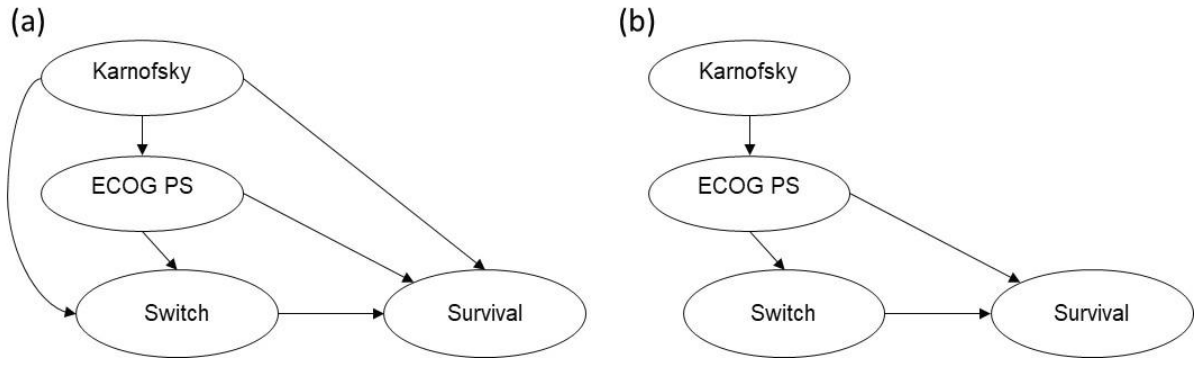
The positivity assumption requires that there are no confounding factors that perfectly predict switching. Hence, the weighting cannot be applied if there are structural differences between the switchers and non-switchers. A relevant structural difference between switchers and non-switchers exists if there is a prognostic characteristic that can determine if a patient is a switcher or non-switcher without information on switch status. For instance, if every patient in a cancer trial who had an Eastern Cooperative Oncology Group (ECOG) performance status score of 1 switched treatment, and if ECOG performance status was a predictor of survival, it would not be possible to apply IPCW to adjust for the switching. In the absence of structural differences, it is still useful to consider this assumption when constructing categorical dummy variables from levels of continuous data or from other categorical variables to avoid creating variables that perfectly predict switch or no switch.[11]

The no unmeasured confounding assumption requires that prognostic variables that influence switch and death are included in the model. The assumption cannot be perfectly tested, but expert clinical opinion and DAGs should be used together to identify which covariates should be included in switching models.[10, 20] There is often concern that too few variables are included in IPCW analyses, making the no unmeasured confounders assumption unlikely to hold. This is important, but it is also important not to include unnecessary variables in models because (i) the inclusion of an inappropriate variable may result in selection bias from collider stratification, (ii) the inclusion of too many variables relative to the sample size may result in finite-sample bias, and (iii) the confidence intervals around point estimates will be larger with the inclusion of non-confounding variables.[11]

Satisfying the no unmeasured confounders assumption should not be taken to mean that *every* prognostic variable must be included. Variables that affect switching but do not affect survival, or that affect survival but not switching, are not needed, i.e. are not

confounders. In addition, only variables that have an independent effect on switching and survival are required – for example, imagine a clinical trial collected information on ECOG performance status but not on Karnofsky score, and that a clinical expert states that both these variables are prognostic for survival and may be related to the switch decision. Failure to collect information on the Karnofsky score only results in unmeasured confounding if it affects switching *and* survival *independent* of the ECOG score – see Figure 1(a) where Karnofsky score is important because there are arrows from Karnofsky score to the switching variable *and* to survival. If the impact of Karnofsky score on switching and survival is fully explained by its relationship with ECOG performance status, then Karnofsky score would not be needed in switching models (see Figure 1(b)).

Figure 1: (a) Simplified DAG where Karnofsky score and ECOG PS are both important confounders; (b) Simplified DAG where only ECOG PS is an important confounder



DAG: Directed Acyclic Graph; ECOG PS: Eastern Cooperative Oncology Group Performance Status

Further, only information that is available to the clinician (or patient) can cause confounding, because if a variable (or the value of a variable) is unknown, it cannot directly affect the switching decision. Therefore, missing data must be dealt with carefully. A variable might be observed and be a confounder for one patient, but may be missing and therefore not a confounder for another patient. Therefore, it may be important to create variables to indicate the missingness of variables, and it may be reasonable to use a last observation carried forward approach in switching models. However, imputation techniques may be important if the value of a variable was likely to have been known by the treating clinician but was not recorded in the trial dataset.

It is important to note that the IPCW method is prone to serious bias when data collection is stopped at a point during the trial (for example at disease progression), when switching could occur some time after that point. This resonates with the ICH E9 addendum referred to in Section 3, which states that trials should be planned in order to ensure that data are collected to support the application of methods required to address *all* estimands of interest.

Often it may be difficult to identify optimal formulations of switching models, with respect to which covariates are included and how they are parameterised. Akaike information criterion (AIC) and the Bayesian information criterion (BIC) may be used to provide information on relative model fits, but the clinical plausibility of the no unmeasured confounding assumption is of greatest importance. The range of weights estimated by different model formulations should also be considered and reported (see Section 6). Versions of models can also include different ways of modelling time (e.g. with cubic splines) and/or categorical variables defined in different ways. Readers are referred to Hernán and Robins (2020), Tennant *et al.* (2021), among others for more information on DAGs and covariate selection.[10, 19, 20, 32]

Truncation of extreme weights

The IPCW method performs better when sample sizes are large and with small or moderate proportions of switchers.[10, 33, 34] In contrast, when switching proportions are high IPCW results can be prone to substantial bias, especially when sample sizes are small, primarily because in these cases extreme weights are common.[10, 33, 34] Some studies have proposed methods to deal with extreme inverse probability weights. Cole and Hernan (2008) describe the trade-off between bias and precision associated with the truncation of weights.[11] Truncation of weights involves selecting a maximum value that the weights can take. The value is selected based upon a pre-specified percentile of the weight distribution. Often the 99^{th} or 95^{th} percentile of the weight distribution is chosen. As truncation of the distribution is increased, the estimates become more biased due to loss of information, but standard errors of the weights and therefore of the estimated treatment effects are reduced.[11] Bai *et al.* (2015) propose an adaptive truncated marginal structural model whereby weights are truncated at each time point to prevent extremely large weights being propagated over time.[35] If weights are truncated, a sensitivity analysis should be undertaken around

the truncation percentiles to assess the robustness of the results to changes in the truncation percentiles.

4.2.3 Two-stage estimation (TSE)

The TSE approach involves first estimating the effect on survival associated with treatment switching, and then using this effect to estimate survival times that would have been observed if switching had not occurred.[5, 36] The method can be used to adjust for treatment switching from the control group to the experimental treatment, and/or from either trial arm to other subsequent treatments. The TSE approach requires that switching only occurs at or after a disease-related “secondary baseline” time-point.[5] Often, disease progression fits the criteria of a suitable secondary baseline. The method also requires the no unmeasured confounding assumption to hold, whereby switching must be independent of potential outcomes, conditional on patient characteristics measured (and included in the model) at the secondary baseline. The simple version of the TSE method described by Latimer *et al.* (2017) also assumes that if switching occurs after the secondary baseline, there is no time-dependent confounding between the secondary baseline time-point and the switch time-point – that is, there should be no important prognostic changes in patients between these time-points.[5] An extension to this simple approach is described in Section 4.2.3.1.

In the situation where patients randomised to the control group switch onto the experimental treatment after disease progression, the application of the simple TSE method involves applying an accelerated failure time (AFT) model to compare post-secondary baseline (i.e. post-progression) survival in control group switchers with post-secondary baseline survival in control group non-switchers. The AFT model includes prognostic variables measured at the secondary baseline time-point to account for differences in prognosis between switchers and non-switchers. The treatment effect associated with switching is estimated in the form of a time ratio, and is used to adjust survival times in switchers using a counterfactual survival model similar to that described in equation (1). Through this process an adjusted dataset is derived. Survival analysis can then be undertaken on the adjusted dataset, to estimate a treatment effect adjusted for treatment switching, and/or to fit parametric models to

extrapolate survival beyond the trial period. As for the RPSFTM, standard or more flexible survival models can be applied to the adjusted dataset (see TSD 21 [21]), but uncertainty should be appropriately characterised by bootstrapping the entire adjustment and subsequent survival model-fitting process.

4.2.3.1 Extensions and recent developments to TSE

Issues relating to re-censoring

The recent research on re-censoring by Latimer *et al.* described in Section 4.2.1.1 is relevant for the TSE method as well as the RPSFTM and IPE methods, and therefore it is recommended that TSE should be applied with and without re-censoring.[25] Further to this, Latimer *et al.* (2019) tested an application of TSE that used IPCW to deal with potentially informative censoring in the TSE-adjusted dataset, as an alternative to re-censoring.[33] This method performed well in simulations and provided results between the two extremes of with and without re-censoring, but involves increased complexity and is subject to the same limitations associated with applying IPCW in general, most notably the assumption of no unmeasured confounding.

Adjusting for multiple types of switching

TSE can be applied to adjust for switches to subsequent treatments in either or both randomised arms. When adjusting for switching in both arms, a separate AFT model can be applied to each arm to compare switchers with non-switchers from the secondary baseline time-point. Survival times for switchers should then be adjusted using the estimated subsequent treatment effect from the relevant treatment arm. If it is relevant to adjust for different types of switching within each treatment arm (for example, if some patients switched to subsequent treatment A, and some switched to subsequent treatment B), potentially an AFT model could be used to estimate the effect of each subsequent treatment separately, or one average effect of subsequent treatment could be estimated. Theoretically, calculating separate subsequent treatment effects may be preferred due to an increased level of detail. However, it is possible that the benefits of such analyses may be minimal, and that they are likely to suffer from small sample sizes, potential error, and uncertainty – further research in this area would be valuable.

Ouwens *et al.* (2021) applied a modified two-stage method (M2SM) using time of switch as a secondary baseline, rather than time of disease progression.[37] The authors analysed a trial for a new treatment for unresectable stage III non-small cell lung cancer, in which patients in both arms of the trial were permitted to receive immunotherapy after discontinuing their study treatment. The authors aimed to adjust for the impact of immunotherapy on overall survival, but decided that they could not use disease progression as the secondary baseline time-point, because in a large proportion of patients switching occurred at a time point substantially after disease progression (median switch time was approximately 6 months after progression), and hence using disease progression as the secondary baseline could have resulted in important time-dependent confounding. The authors state that this time-dependent confounding could be eliminated by using the time of switch as the secondary baseline.

In general, using switch time as the secondary baseline in a TSE analysis is problematic; firstly because patients who do not switch must be excluded from the analysis (even if they have discontinued study treatment and are following a standard treatment pathway – i.e. no subsequent treatment), and secondly because if the secondary baseline time-point does not represent a common disease-related time-point, patients may differ significantly at their respective switch times. Ouwens *et al.* dealt with the first issue by estimating the effect of subsequent immunotherapy treatment compared to subsequent 'other' treatment – i.e. assuming that if a patient had not switched onto immunotherapy they would have switched onto some other treatment (rather than no treatment at all). They dealt with the second issue by using prognostic information measured at (or before) the time of switch to control for differences between patients who switched onto immunotherapy and patients who switched onto other treatments.

Ouwens *et al.*'s M2SM represents a useful adaptation to the simple TSE method, and may be preferable to it when there are large time intervals between disease progression and the time of switch. However, concerns around the lack of a common disease-related secondary baseline remain, and, potentially, a method that uses a common disease-related baseline and is able to deal with time-dependent confounding may be preferred. This may be achieved using TSE with g-estimation.

TSE with g-estimation

Latimer *et al.* (2020) describe an extension to the simple TSE method, which takes the form of a structural nested model (SNM) and uses g-estimation, referred to as TSEgest.[10] The TSEgest method replaces the simple AFT model used in the simple TSE method with g-estimation and a SNM, so that time-dependent confounding can be accounted for. Therefore, switching that is delayed for a period of time after the chosen secondary baseline is no longer a problem for the method, provided that information on time-varying characteristics that predict switching and survival is measured beyond the secondary baseline time-point.

The approach involves a model for switching, and a model for counterfactual survival times. The model for switching has a binary switch-dependent variable which represents observed switch status for patient *i*. Explanatory variables include all confounding variables for patient *i* measured at baseline and over time, and counterfactual survival time for patient *i* from the secondary baseline. The counterfactual survival time model is similar to that represented by equation (1). The switching model and counterfactual model are used simultaneously to obtain the g-estimate of ψ. A value is chosen for ψ, and the counterfactual survival time associated with this value is estimated for each patient. This counterfactual survival time is then substituted into the switching model, and the "true" estimate of ψ is the one that results in the coefficient of the counterfactual survival time variable in the switching model equalling zero – indicating that switching is independent of counterfactual survival time, conditional on all the other variables included in the switching model.

Once identified, ψ can be used to generate the post-secondary baseline survival times adjusted for switching, and adjusted overall survival estimates can be derived. As for the simple TSE method, TSEgest should be applied with and without re-censoring, and could also be applied with IPCW instead of re-censoring.[33] Confidence intervals should be obtained by bootstrapping the entire adjustment process (including estimation of ψ) to adequately capture uncertainty. Similar to the simple TSE method, TSEgest could be used to adjust for switches from either randomised group onto any other subsequent treatments. Switching and outcomes models would need to be specified to compare outcomes between patients who did and did not switch onto the

treatments that need to be adjusted for, allowing the treatment effects of subsequent treatments to be estimated, and adjusted survival times derived.

The key advantage of TSEgest over simple TSE is that it does not rely on the assumption that there is no time-dependent confounding between the secondary baseline and the time of switch. In addition, theoretically the method could also be used when switching occurs before or after the secondary baseline time-point – that is, a secondary baseline is not actually needed – if switching could occur before disease progression, the time of randomisation could be used as the analysis baseline. Such an analysis would require that there is no unmeasured confounding between the time of trial randomisation and the time of switch (for switchers), and between the time of randomisation and the end of trial follow-up (for non-switchers). This is similar to the assumption that would be made by IPCW in the context of switching that is permitted before disease progression, but may represent a stronger assumption than that of assuming no unmeasured confounding between the time of secondary baseline and the time of switch (or end of follow-up) – which is what is required when switching is only permitted after a specified secondary baseline. In general, for both TSEgest and IPCW, the longer the time period between the analysis baseline and the switch time-points, the stronger the assumption of no unmeasured confounding between these time-points.

Latimer *et al.* (2020) demonstrate that TSEgest produces much lower bias than the simple TSE method when time-dependent confounding occurs between the secondary baseline and switch time-points, and that TSEgest is less prone to bias than IPCW in situations with high switching proportions.[10] Hence, the authors conclude that TSEgest represents a more flexible alternative to simple TSE, that can provide additional information to policy makers when presented alongside IPCW, RPSFTM and IPE results. TSEgest does not make the simple TSE method obsolete, but when time-dependent confounding is suspected the simple TSE method should not be relied upon.

Variable selection and directed acyclic graphs (DAGs)

Like the IPCW method, the TSE method (and adaptations/extensions of the method) are reliant upon the no unmeasured confounding assumption. Therefore, the

discussion contained in Section 4.2.2.1, around variable selection, model selection and specification, missing data and DAGs, is as relevant for TSE (and TSEgest) as it is for IPCW – we refer readers back to that section.

4.2.4 Other methodological developments

A random forest-based prediction approach

Xu *et al.* (2021) proposed a random forest-based prediction approach for adjusting for treatment switching.[7] The method uses a random forest algorithm to predict counterfactual survival times to estimate what would have happened in the absence of switching. It requires data on baseline and time-varying prognostic covariates that could influence survival time *and* the probability of switching. It is applied by splitting the data into a training set of non-switchers in the control arm and a prediction set of switchers. These subgroups are split into further subgroups depending upon whether they had an event or were censored. The counterfactual survival times for switchers with an event are generated using the random forest algorithm applied to the training set of non-switchers who experienced an event, and the counterfactual survival times for switchers who were censored are generated using the random forest algorithm applied to the training set of non-switchers who were censored. The random forest algorithm involves taking bootstrap samples from the training set, generating a prediction of the counterfactual survival for each bootstrap sample, and then averaging over the bootstrapped samples. Once counterfactual event times are predicted, re-censoring can be applied and the final survival analysis, adjusted for switching, can be performed.

An advantage of the method is that it does not require a pre-specified model structure to define the relationships between covariates, treatment indicator and outcome. The authors claim that the model does not rely on the no unmeasured confounders assumption, however the model requires data on covariates that predict the counterfactual survival time and the probability of switching, and therefore in practice the data requirements and assumptions for the model are similar to IPCW or TSE. The method is capable of adjusting for switching to multiple treatments, and for switching in both arms of the trial. It is recommended that the number of treatment effects should be kept to a minimum, because computational burden increases with the number of treatment effects included in the model. Treatments which are expected to have similar

effects should be grouped together. According to the results of a simulation study by the authors, the model performed well compared with RPSFTM, IPE, IPCW and simple censoring. Further research is required to test the performance of the approach in a wider range of scenarios.

A regression imputation approach

Luo *et al.* (2016) propose an approach for adjusting for subsequent therapies which extends existing approaches.[38] The authors devise an approach that involves three adjustment factors: for the time at which treatment switching occurs, the treatment effect associated with the treatment switched to, and the treatment group to which the patient was originally randomised. Therefore, while switches to more than one different treatment are not explicitly modelled, the overall impact of the switch can vary according to when the switch occurred and which randomised group the patient was originally allocated to. To estimate the three adjustment factors, two survival models are used. The first relates counterfactual (without switching) survival times to observed survival times as a function of whether a switch occurred, and the time of the switch and which group the patient was randomised to (if a switch did occur). Starting values for the coefficients of each of these three factors are chosen, and counterfactual survival times are estimated. These are then used in a second survival model, where the counterfactual survival times are related to randomised group, baseline covariates, and a function of the three adjustment factors. In this model, the coefficients of the three adjustment factors should be close to zero, since in this model counterfactual (without switching) survival times are being modelled. If these coefficient values are not close to zero, they are taken as the next set of values to substitute into the first survival model. The process is continued until the values of the three coefficients are close to zero.

This iterative estimation process is similar to that used in the IPE method, but, unlike the IPE and RPSFTM, the approach does not rely upon the common treatment effect assumption (as the overall impact of switching is permitted to vary according to three factors), and does not use randomisation to allow the effects of switching to be identified. Instead it is assumed that the three modelled adjustment factors are sufficient for representing the impact of switching and the cause of switching, and that there is no unmeasured confounding in the second survival model.

The authors demonstrate the use of their method in a simulation study where performance is good, but the scenarios considered are relatively simple and the assumptions associated with their method are not discussed in detail. While the method potentially offers more flexibility than the RPSFTM and IPE methods, it does not benefit from being randomisation-based, and therefore requires the no unmeasured confounding assumption. The modelling approach is perhaps less intuitive than the switching and outcomes modelling approach associated with IPCW and more complex TSE methods, and the relationship between prognostic characteristics and switching is not explicitly modelled. Further research on the performance of the method in more realistic settings would be valuable.

A semi-competing risks model

Chen *et al.* (2020) develop a method that can deal with switches to subsequent therapies that extends an approach originally suggested by Zeng *et al.* (2012).[27, 39] The method is also similar to an approach described by Zhang *et al.* (2021) which uses semi-competing risks models applied in a Bayesian framework to estimate treatment effects adjusted for treatment switching.[40]

A semi-parametric mixture model is used that allows the effects of subsequent treatments to differ from that of the initially randomised treatment (therefore the common treatment effect assumption is not required), and also allows the effect of subsequent treatments to differ depending on whether switching happens before or after disease progression, and depending on which randomised group a patient is originally in. This is achieved by using a model with three components and a transition model structure; i) a model for disease progression status, as a function of baseline covariates and randomised group; ii) a model for survival for those who do not progress as a function of randomised group, switching status (and the interaction of switching status and original randomised group), and baseline covariates; iii) a model for time to disease progression and time from progression to death (in those who experience disease progression) as a function of randomised group, baseline covariates (and prognostic information measured at disease progression), and switch status. Patients are then split into four groups, according to whether or not disease progression and death were observed or censored (i.e. yes, yes; yes, no; no, yes; no,

no), and the contribution of each group to the likelihood function is calculated based upon the previously described model components. The proposed models are then used within a counterfactual outcomes framework to estimate survival functions conditional on covariates, randomised group, and switching, such that the potential survival function with no switching can be estimated. The method involves modelling the disease process and is reliant on the no unmeasured confounding assumption.[27]

Chen *et al.*'s method only uses prognostic data measured at baseline and the time of disease progression – the authors state that this could be extended to include factors measured at other timepoints, but it is unclear whether this would cause issues around standard regression and time-dependent confounding. The authors compared their approach to that of Luo *et al.* (2016) (see "a regression imputation approach", above) and showed that Luo *et al.*'s approach produced substantial bias when simulated data was generated in a way that reflected the specification required by Chen *et al.*'s approach, whereas Chen *et al*.'s method performed reasonably well when tested using Luo *et al.*'s simulation mechanism.[27, 38] Comparisons to other adjustment methods in more general settings were not made, and therefore further research would be valuable.

A semi-parametric copula-based model

Huang *et al.* (2020) describe a semi-parametric copula-based model to estimate treatment effects in the presence of treatment switching.[41] The method is designed to deal with switching that happens immediately upon disease progression. The model described by the authors only includes one type of treatment switch, but the authors state that this could be extended. The approach consists of a copula model for the joint distribution of time-to-progression and overall survival (which includes terms for randomised group and baseline covariates), and a conditional hazard model for overall survival subsequent to disease progression, which includes terms for randomised group, baseline covariates, covariates measured at the time of disease progression, and treatment switching. The coefficients of the randomised group and treatment switching terms are interpreted as the effects of randomised treatment (with no switching effect), and the effect of switching, respectively. Because the authors use a proportional hazards framework, treatment effects on the hazard scale are estimated. The authors state that time-dependent covariates could be accommodated (rather

than simply covariates measured at baseline and at the time of disease progression), but it is unclear how potential time-dependent confounding would be dealt with.

Huang *et al*.'s approach has similarities to the TSE method, in that a post-progression effect of switching is estimated which in turn allows the effect of randomised treatment to be identified. However, the method does not readily generate counterfactual survival times, and so it may be less useful in an HTA context. The method also includes the limiting assumption that switching happens immediately upon disease progression, and so would not use actual switching times if these occurred after disease progression – this may result in face-validity issues. The authors state that the method does not require the no unmeasured confounders assumption, but in fact analyses will depend upon inclusion of sufficient covariates to allow valid treatment effects to be estimated, and therefore we believe that the no unmeasured confounding assumption is still implicitly required.

A decision-analytic modelling approach

Kuehne *et al.* (2021) describe a novel decision-analytic modelling approach for estimating survival outcomes that would have been observed in a clinical trial if treatment switching had not occurred.[42] Unlike all the other methods described in this document, patient-level data from the trial affected by treatment switching is not required.

The authors describe two trials that investigated the use of bevacizumab to treat ovarian cancer. In one trial (ICON7) treatment switching from the control group onto bevacizumab was not permitted, and significant progression-free and overall survival benefits in favour of bevacizumab were shown. In the second trial (GOG-218), 70% of patients randomised to the control treatment switched onto bevacizumab after disease progression and a significant benefit was shown for progression-free survival, but not for OS. The authors developed a "causal decision-analytic Markov model" (CDAMM) which they used to emulate the GOG-218 trial. The two arms of the GOG-218 trial were emulated, and the modelled population was defined by age and disease status according to summary data on the characteristics of the trial population. The model structure was based upon a DAG, with the presence of malignant ascites thought to represent a key prognostic characteristic that was affected by treatment with

bevacizumab, affected the probability of a control group patient switching onto bevacizumab, and affected the rate of mortality after disease progression. Summary data from a range of RCTs (including GOG-218) and other published literature were used to inform transition probabilities in the model. Once the model was constructed, the authors calibrated its output to match the GOG-218 Kaplan-Meier overall survival curves by modelling the 70% switch proportion and varying the characteristics of switchers with respect to presence of malignant ascites. Once calibration had been achieved, the switching proportion was set to 0% to estimate outcomes that would have been observed with no switching.

Kuehne *et al.*'s method is novel, in that it attempts to estimate outcomes adjusted for treatment switching in the case where patient-level data (and therefore information on key prognostic characteristics and their relationship with the probability of switching) are not available. Work on alternative methods to adjust for treatment switching when only summary data are available have previously been presented as research abstracts,[43, 44] but full peer-reviewed journal articles have not followed. Such approaches would make the application of adjustment analyses possible for a wider range of analysts – in the context of NICE appraisals, it is very rare for external assessment groups to have access to patient-level data from the RCTs under investigation. In their example, Kuehne *et al.* show that their switching-adjusted results matched those from the ICON7 trial (which investigated bevacizumab for ovarian cancer and was not affected by treatment switching) reasonably closely. However, the authors acknowledge that their approach cannot model the switching process in detail, and is reliant on several important assumptions around the treatment effect in switchers, the characteristics of switchers, and the disease process. Whilst this novel approach represents a potentially valuable area of research, approaches that attempt to adjust for treatment switching without making use of patient-level data from the trial in question are likely to be regarded as more speculative than those that are able to leverage the trial data.

Using external data

The focus of TSD 16 and of this update is on methods that adjust for treatment switching by making adjustments to observed data from the pivotal trial in which treatment switching occurred. Alternative approaches could involve the use of external

data. Kuehne *et al.*'s method makes use of a wide range of data, including data from external trials,[42] and Li *et al.* suggest supplementing their "enhanced" RPSFTM with the use of the literature and expert opinion when attempting to adjust for several treatments with different effects.[8] A more direct use of external data was identified in TSD 16: in the NICE appraisal of lenalidomide for multiple myeloma, in which the pivotal trial was affected by patients in the control group switching onto lenalidomide, an analysis was presented that used patient-level data from previous trials of the comparator treatment to estimate control group survival.[45, 46] This type of analysis – whereby external data sources are used to estimate comparator group survival outcomes due to the confounded control group data observed in the pivotal clinical trial – is akin to methods used to estimate comparative effectiveness in uncontrolled (or single arm) studies. We did not identify any new methods that relied solely on using external data to estimate counterfactual survival times for switchers (or for entire treatment arms affected by switching). However, in recent years a substantial amount has been written about estimating comparative effectiveness from uncontrolled studies and, relatedly, about analysing observational (or "real world" data) – TSD 17 focuses on the analysis of observational data, TSD 18 describes methods for conducting population-adjusted indirect comparisons, and the NICE real world evidence framework covers related issues, including the use of external control studies.[47-49]

Using external data sources to estimate counterfactual survival times for switchers (or for entire treatment arms affected by switching) would require access to patient-level data from the external data source, and would require there to be no unmeasured confounding such that any prognostic differences between the patients included in the pivotal trial and those included in the external data could be controlled for – for example, external control data could be matched or adjusted to represent control group patients who switched treatment in the pivotal trial. If external data were used to completely re-estimate survival times for an entire trial arm affected by switching, a large amount of data from the pivotal trial would be discarded, including data for non-switchers and data observed in switchers before switching occurred. For this reason, it is likely to be preferable to use the adjustment methods described in this TSD, which do not discard these data. However, an alternative, and possibly more attractive, approach may be to attempt to use external data to estimate counterfactual survival

beyond the switching time-point, resulting in an adjusted dataset that includes data from the pivotal trial for non-switchers and for switchers up to the switch time-point, and data derived from the external data source for the post-switch period in switching patients. We are not aware of any papers that describe such a method – further research may be valuable, particularly given the increasing availability of routine electronic health record (EHR) data (e.g. SAIL Databank for the Welsh population and Systemic Anti-Cancer Therapy [SACT] data for England) and the fact that these data sources will often represent outcomes associated with standard treatments in a particular jurisdiction.

We believe that in most cases it will be preferable to base adjustments for treatment switching on data observed in the pivotal trial in which switching took place. However, in some instances it may be appropriate to attempt to use external data to estimate survival times that would have been observed in the absence of treatment switching. In such cases, methods should follow those recommended in relevant guidance documents related to the use of matched controls and analysing observational data, including TSDs 17 and 18, and the NICE real world evidence framework.[47-49]

5. TREATMENT SWITCHING IN NICE APPRAISALS

5.1 INTRODUCTION

In this section, we provide a review of treatment switching in recent NICE TAs. The aim of this review is to identify instances where treatment switching was discussed, ascertain the circumstances of the treatment switching, establish which methods were used to address treatment switching, how applications of these methods were presented in submissions to NICE, and highlight issues that have arisen in practical applications of treatment switching adjustment methods. A similar review was undertaken in TSD 16 for the period 2000-2009. Since publication of TSD 16, we expected to see greater use of more complex treatment switching adjustment methods than observed previously.[1]

5.2 METHODS

We identified TAs of cancer treatments published between 1st January 2020 and 20th April 2022 (accessed 20th April 2022) from the NICE website. We focused on cancer TAs because switching is most prevalent in these, but we acknowledge that switching can also be important in TAs in other disease areas. Searches were performed to find the terms “switch” and/or “crossover” within the Final Appraisal Document (FAD) for each of the cancer TAs (noting that the appraisals we reviewed were initiated before the term ‘FAD’ was replaced with ‘Final draft guidance’). An assessment of the surrounding text was made to establish if the terms “switch” or “crossover” referred to switching or crossover in a trial. For those TAs that referred to switch or crossover in a trial, a review was performed of the FAD, evidence review group (ERG) report (noting that the appraisals we reviewed were initiated before the term ‘ERG’ was replaced with ‘External Assessment Group’), company submission (CS) and where relevant, other publicly available documents on the NICE website. From the review we collected data on the context in which the switches occurred, the adjustment methods applied, and the reporting associated with the application of the method.

5.3 RESULTS

In the 40 cancer TAs reviewed, 10 of the FADs mentioned switching or crossover. These were the FADs for TA784, TA742, TA741, TA740, TA709, TA705, TA670, TA660, TA653 and TA643. Within these TAs, treatment switching adjustment methods

were applied to the trial data in 7 TAs.[50-56] In 3 TAs,[57-59] treatment switching adjustment methods were not applied to the pivotal trials themselves, however treatment switching adjusted external trials were used in indirect treatment comparisons (ITC) to inform the appraisal. Adjusted summary statistics were used in the ITC without any in-depth consideration recorded in the FAD or CS of appropriateness of the adjustment methods used to create the summary statistics.

Of the 7 TAs with switching in the pivotal clinical trial, 4 were affected by switching from the control arm to the experimental treatment, and 6 were affected by switching from either arm to subsequent therapies that were not available in England and Wales at the time of the submission. 3 involved trials that were affected by both types of switching. Table 1 provides a summary of the type of switching and the treatment switching adjustment methods used in the TAs. None of the TAs applied all four of the complex methods suggested in TSD 16. 3 TAs applied more than one adjustment method.

Table 1: Treatment switching in NICE TAs

TA number	Switching from control to experimental	Switching to subsequent treatments	Used RPSFTM	Used IPE	Used IPCW	Used TSE	Used another switching adjustment method
TA784		X			X		
TA740	X	X					X
TA741		X			X		X
TA709	X	X	X		X	X	
TA705		X	X				
TA660	X	X	X	X			
TA653	X		X				

TA: Technology Appraisal; RPSFTM: rank preserving structural failure time model; IPE: iterative parameter estimation; IPCW: inverse probability of censoring weights; TSE: two-stage estimation.

Within these 7 TAs, 5 state that additional information on the switching adjustment analyses are included in submission appendices, which are not publicly available. Only one TA [53] contained a discussion of the key methodological assumptions in relation to the trial within the main company submission document, and a further 3 [51, 52, 56], provided some justification of the methods used. Of the 6 TAs that applied RPSFTM, IPE, TSE or another method that was compatible with re-censoring, only 4 TAs mentioned re-censoring in the company submission.[51-53, 56] It was not clear how re-censoring had been dealt with in the other TAs.[54, 55]

TA784 was a Cancer Drugs Fund review of TA528 (niraparib for maintenance treatment of relapsed, platinum-sensitive ovarian, fallopian tube and peritoneal cancer).[50, 60] The NOVA trial (niraparib vs placebo) had two cohorts: participants with germline breast cancer gene mutated (gBRCAm), and participants with non-gBRCAm. The trial was confounded by subsequent switches to poly adenosine diphosphate-ribose polymerase (PARP) inhibitors in both trial arms, and an issue in the appraisal was that PARP inhibitors are not available as the next line of treatment in standard clinical practice in England and Wales. In the non-gBRCAm cohort, 234 were randomised to niraparib and 15 (6.4%) of those switched onto a subsequent PARP inhibitor, and 116 were randomised to placebo and 15 (12.9%) of those switched. IPCW was applied to adjust for switching to PARP inhibitors. Estimated treatment effects and Kaplan-Meier survival curves were very similar between adjusted and non-adjusted analyses. There was no information in the publicly available documents submitted to NICE on how the IPCW analysis was conducted, the range of the calculated weights and whether these were stabilised, or which covariates were included in the switching and outcome models and how these were selected. Nor was there any indication that information on these aspects were provided in an appendix – indeed, in its report, the ERG stated that they could not comment on the robustness of the IPCW analyses due to the sparse information provided.[61]

TA740 appraised apalutamide with androgen deprivation therapy (ADT) for treating high-risk hormone-relapsed non-metastatic prostate cancer, and TA741 appraised apalutamide with ADT for treating hormone-sensitive metastatic prostate cancer.[51, 52] The pivotal trial for TA740 was SPARTAN, and for TA741 was TITAN, both of which compared apalutamide plus ADT to placebo plus ADT. In the SPARTAN trial, switches were made from the control arm to the experimental treatment and from both trial arms to ‘other’ subsequent treatments, such as abiraterone and enzalutamide. The Appraisal Committee noted that standard practice in England and Wales is for people to have only one androgen receptor inhibitor (such as abiraterone, enzalutamide, or apalutamide), and this was taken into account when deciding which treatment switches it was relevant to adjust for. The Committee also recognised that in some circumstances it may be reasonable to assume that drugs of the same class have similar effectiveness, and so the RPSFTM method could be appropriate even if

switches are onto ‘other’ treatments. The company used a modified RPSFTM approach to adjust for treatment switching, a currently unpublished method which was proposed by Diels *et al.* (2019) in a poster presented at the International Society for Pharmacoeconomics and Outcomes Research (ISPOR) conference.[62] The justification for using this method was that the standard single-parameter RPSFTM cannot deal with switches to ‘other’ subsequent treatments. The company attempted to apply IPCW but stated that it "generated counter-intuitive and clinically implausible results" (TA740, Draft guidance 1, Committee papers, pp102 [52]). The modified RPSFTM approach involved a hybrid of the TSE and RPSFTM methods, where the treatment effect associated with ‘other’ subsequent treatments was estimated based upon analyses of the effects of these treatments in an external trial, allowing counterfactual survival times to be derived for those in SPARTAN who switched onto ‘other’ treatments. An RPSFTM analysis was then used to adjust for participants who switched from the control group onto apalutamide. This bears some similarity to the Li *et al.* (2017) approach summarised in Section 4.2.1.1.[8] In their analysis of the TITAN trial, relevant for TA741, the company applied the same modified RPSFTM as well as IPCW to adjust for switches to a second androgen receptor therapy in both arms of the trial. The company found that there was no significant difference between the adjusted and unadjusted analyses, which may have been due to the relatively small number of switchers. In both TA740 and TA741 the details of the adjustment analyses were provided in submission appendices, which are not published on the NICE website, so we are unable to describe the information provided regarding the methods used. The publicly available appraisal documents do not include a substantial amount of technical detail on the switching adjustment analyses, instead providing a general description of how the modified RPSFTM was applied, a discussion of the key assumptions of the methods, and the results of the analyses presented as HRs.[51, 52]

In TA709, the pivotal clinical trial was KEYNOTE-177, which compared pembrolizumab to chemotherapy for untreated metastatic colorectal cancer with high microsatellite instability (MSI-H) or mismatch repair deficiency (dMMR).[53] 56 participants (36%) randomised to the control arm switched onto pembrolizumab and a further 35 control arm participants switched to other anti-PD-1/PD-L1 (Programmed Cell Death Protein 1 / Programmed Cell Death Ligand 1) therapies, which are not

available at second line and beyond in standard clinical practice in England and Wales. The company applied RPSFTM, the simple TSE method and IPCW to adjust for both types of switching in the control arm. Of the switching adjusted analyses, the company favoured their application of the TSE method without re-censoring, which they used as a sensitivity analysis to the primary ITT-based OS analysis. Justification of their favoured method was provided in the main text of their submission by discussing the assumptions of each of the adjustment methods applied in the context of the KEYNOTE-177 trial. The RPSFTM was disregarded because the company believed the common treatment effect assumption was unlikely to hold, and assessment of the data supported this claim. The IPCW approach was also disregarded because data had not been collected on all relevant confounders, therefore the no unmeasured confounders assumption was not considered to be valid. TSE was preferred because it does not rely on the common treatment effect assumption. However, it was not mentioned that TSE relies on adequate data collection at the secondary baseline (disease progression in this case), and requires that switching happens at or shortly after the secondary baseline, which did not appear to hold for all switchers in the trial. The committee papers indicate that switching was not only from the control group onto pembrolizumab – patients in both trial arms also switched to other subsequent immune checkpoint inhibitors, and some switching occurred before disease progression (5 in the pembrolizumab arm and 15 in the standard care arm). In the primary ITT analysis of progression-free survival, these patients were simply censored at initiation of their new-anti-cancer treatment. This simple censoring approach is prone to bias. There was no comment on the appropriateness of this censoring approach in the FAD. In addition, whilst the company appeared to conduct analyses to adjust survival estimates in the control arm for switches to pembrolizumab and to other anti-PD-1/PD-L1 therapies, there is no mention of any adjustments made for switching in the experimental arm of the trial. Failing to adjust for switches to other anti-PD-1/PD-L1 therapies in both arms will result in bias in favour of pembrolizumab if other anti-PD-1/PD-L1 therapies are beneficial treatments and if those treatments do not represent standard treatment pathways available in England and Wales. This issue was not discussed in the FAD or company submission.

In TA705, the pivotal trial compared atezolizumab to platinum based chemotherapy (combined with pemetrexed or gemcitabine) for the treatment of stage IV non-small

cell lung cancer (NSCLC).[54] Switching was not permitted in the trial, but some participants received subsequent lines of cancer therapies, including non-protocol immunotherapies. The company applied the RPSFTM as a sensitivity analysis, to estimate what the treatment effect might have been if switches onto subsequent lines of immunotherapies had not occurred. The details of this analysis were provided in an appendix to the submission document which is not in the public domain, therefore we cannot assess the validity of the assumptions imposed. If a standard RPSFTM was applied, it seems likely to have assumed that the treatment effect of the non-protocol immunotherapies that patients switched to would be similar in magnitude to the treatment effect of atezolizumab. The FAD does not discuss the RPSFTM analysis, or report any discussion on whether it was appropriate to adjust for treatment switching in this case.

In TA660, the pivotal clinical trial was ARAMIS, which compared darolutamide plus ADT to placebo plus ADT for the treatment of men with non-metastatic castrate resistant prostate cancer with high-risk of developing metastatic disease.[55] In the trial, 31% of patients randomised to the control group switched onto darolutamide. The company adjusted for this switching using the IPE and RPSFTM methods. These analyses were presented as sensitivity analyses but the company did not present details of how they were applied, and did not explicitly discuss their assumptions. No appendix containing this information was referred to. Results for the switching adjusted analyses were presented as estimates of predicted survival percentages at 5, 10, 15, 20 and 25 years by arm, using parametric survival models fitted to the adjusted data. Switches to other subsequent therapies that are not recommended for use within England and Wales also occurred in ARAMIS, but no adjustments were made for these. The company stated that it did not adjust for these switches because "First, it is not clear how one would combine the crossover adjustment for switching from control to intervention with adjustment for non-UK standard subsequent treatment use, as these are not mutually exclusive given that some patients could have both events in a sequence. Second, it may require pooling treatments to make this viable with the low number of patients upon stratification, and as such this would result in further challenges regarding the assumption of a pooled treatment, which would introduce further uncertainties into the long-term treatment effect." (TA660, Final draft guidance 1, Committee papers, Company response to NICE’s request for clarification,

pp224.[55]). In Section 4.2.1.1 we note that the RPSFTM adjustment method is unlikely to be appropriate when there is a need to adjust for multiple treatment switches, whereas both IPCW and TSE can deal with these situations (see Sections 4.2.2.1 and 4.2.3.1). However, we also note that this remains an area for further research (see Section 7.1).

In TA653, the pivotal clinical trial was AURA3, which compared osimertinib to platinum-based doublet chemotherapy for patients with a confirmed diagnosis of advanced or metastatic (Stage IIIB-IV) epidermal growth factor receptor (EGFR) T790M non-small cell lung cancer, who have progressed following prior therapy with an approved EGFR tyrosine kinase inhibitor (TKI) agent.[56] 100 patients (71%) in the control arm switched to osimertinib. The company chose RPSFTM as its preferred adjustment method, stating that IPCW and TSE are known to produce unreliable results when the proportion of switchers is high. The company presented 6 applications of the RPSFTM – “on treatment” and “treatment group” RPSFTM analyses were applied with re-censoring; without re-censoring; and with re-censoring of the acceleration factor. From the company submission, it is unclear how re-censoring of the acceleration factor was applied. Further details on the application of the methods were included in an appendix, which is not publicly available.

5.4 SUMMARY

NICE TAs of cancer treatments were reviewed for the period January 2020 to 20th April 2022. 7 TAs were identified in which treatment switching was an issue in the pivotal trial. Switches were reported to occur from control arm to experimental treatment (n=4), or to subsequent treatments not available in standard practice in England and Wales (n=6). There were some discussions regarding which switches were appropriate to adjust for in company submissions and recorded in FADs, but this was not done systematically. In TA660, the company chose not to adjust for subsequent therapies that are not available in England and Wales, due to complexities associated with the application of an adjustment analysis in the context of two types of switching. In TA741 and TA740, adjustments were made for both types of switching.[51, 52] In this TSD we have described how methods can be used to adjust for more than one type of switching in the same trial.

A range of switching adjustment methods were applied across the TAs identified, including RPSFTM, IPE, IPCW and TSE. However, none of the TAs applied all of the methods suggested in TSD 16 and some applied alternative methods that were not described in TSD 16.[1] TA740 and TA741 applied a modified RPSFTM approach (Diels *et al.* 2019).[51, 52, 62] Using alternative methods is acceptable, provided that these are appropriately described and justified. It will not always be necessary to apply all of the adjustment methods suggested in TSD 16. However, the rationale for excluding or including methods should be carefully justified – see Recommendation 7 in Section 4.1, and Recommendation 8 in Section 6.

When faced with a situation where different types of treatment switching occur in a trial – for instance, where participants randomised to the control group switch onto the experimental treatment, and participants randomised to either group switch onto some other treatments – applications of adjustment methods typically require additional decisions and assumptions to be made. Firstly, as discussed in Section 2, it is necessary to determine which switches it is appropriate to adjust for, related to whether or not the observed switching is representative of treatment pathways available in standard clinical practice in the jurisdiction for which an analysis is being undertaken. In Sections 4.2.2.1 and 4.2.3.1 we explain that IPCW and TSE methods are able to adjust for multiple switches, when required. In addition, the company submission in TA660 highlights complications around individual patients making more than one switch sequentially, and the need for pooling together treatments to apply methods more simplistically.[55] Neither of these issues mean that a treatment switching adjustment analysis cannot be done. For sequential switches, adjustment can be made from the point of the first switch to a non-standard treatment. For pooling, expert advice should be sought to enable the grouping together of treatments with similar treatment effects. Again, either IPCW or TSE methods could be used - although we note that this remains an area for further research (see Section 7.1).

Justification for the adjustment methodology used was provided in the company submissions of 4 of the TAs.[51-53, 56] The no unmeasured confounding and common treatment effect assumptions were directly referred to in the main text of only one company submission (TA709). In TA740, TA741 and TA653, the common treatment effect assumption was mentioned in the ERG report,[51, 52, 56] and the no

unmeasured confounding assumption was also mentioned in the ERG reports for TA740 and TA741.[51, 52]

None of the TAs provided information on how covariates had been selected for applications of IPCW or TSE. In Section 4.2.2.1 we explain how DAGs can be useful for this purpose, and we recommend that covariate selection is addressed in reporting (see Recommendation 8, Section 6). Re-censoring was referred to in the company submissions of 4 of the TAs,[51-53, 56] and analyses were presented with and without re-censoring in the company submission of 3 TAs.[51, 52, 59] Key diagnostics for the models, such as the range of weights for IPCW, or g-estimation plots for RPSFTM, were not presented in the main text of any company submissions. In at least 5 of the TAs,[51-53, 56] important information on the application of the adjustment methods was included in appendices to the company submission, but these are not publicly available. It was unclear whether additional information on the application of the treatment switching methods was available in unpublished appendices for TA784 and TA660.[50, 55] Given the importance of these analyses, we suggest that the information on the application of the methods should be included in publicly available documents.

6. RECOMMENDATIONS AND REPORTING GUIDELINES

In Section 2 we provided an over-arching recommendation (Recommendation 1) stating that it may be appropriate to attempt to adjust for various different types of switching observed in an RCT in order to address the HTA decision problem. The switching observed in each treatment arm of a trial should always be described and it should be considered whether the observed switching is representative of treatment pathways available in standard clinical practice. However, we note that determining that it is relevant to adjust for a particular type of treatment switching does not mean that an adjustment analysis will be accepted, due to uncertainties around the reliability and potential bias associated with adjustment analyses.

In Sections 3 and 4.1 we provided recommendations for the planning of trials when treatment switching is anticipated (Recommendations 2-7). We refer readers back to those sections, and in particular re-iterate that when planning trials and pre-specifying analyses, trial sponsors should consider questions of interest not only for regulatory agencies, but also for HTA agencies, and especially the implications that this can have for data collection.

Our review of recent NICE TAs, presented in Section 5, highlights key issues associated with treatment switching and adjustment analyses in NICE appraisals. It is clear that the reporting of analyses that adjust for treatment switching, and the review of those analyses, has been sub-optimal. To encourage clear, transparent and consistent reporting, we present a set of reporting guidelines to be followed when submitting analyses that adjust for treatment switching in submissions made to NICE (see Recommendation 8, below, which includes all our reporting guidelines). Sullivan *et al.* (2020) provided a set of reporting guidelines for IPCW and RPSFTM to support improved interpretation of analyses undertaken to adjust for switching.[12] We have added to their guidelines and extended them to cover IPE and the TSE adjustment methods. For clarity, our additions and extensions are highlighted with an asterisk. In some cases, we have altered the wording of the recommendations made by Sullivan *et al.* for clarity. For further context on the application and reporting of IPCW and RPSFTM, and for a review of the reporting of these methods in NICE TAs and journal articles published prior to 22 May 2018, we refer readers to Sullivan *et al.* (2020).[12]

The reporting guidelines cover a description of the types of treatment switching observed, how adjustment methods were chosen, and detailed description of the assumptions and output of the adjustment methods used. Providing this information on a case-by-case basis should ensure that adjustment analyses used in NICE appraisals are consistently and rigorously justified and reported.

Recommendation 8: *Reporting guidelines for treatment switching adjustment analyses*

Description of the data and unadjusted results

D1. Provide unadjusted results from an ITT analysis for comparison.

D2. Describe the treatment switching mechanism - who can switch and when.

D3. Detail the number of patients who switched, the number eligible to switch and when switching occurred.

D4. Give an overview of the data available for adjustment - what predictors were collected and how frequently were they measured.

D5. Include a summary of subsequent treatments received in both/all arms of the trial, including which subsequent treatments were received, the number and proportion of patients that received subsequent treatments, and when subsequent treatments were received.*

D6. Describe which switches do not represent standard treatment pathways in the NHS in England and Wales.*

D7. Describe and justify the type of switches that adjustments have been made for.*

Method selection

M1. State whether the chosen adjustment approach, including all model fitting steps, was prespecified; if not, explain how the final method and model was selected.

IPCW

IPCW1. Provide a statement around the plausibility of the no unmeasured confounders assumption (this can include a relevant DAG for selection of variables).*

IPCW2. Provide summary statistics for each variable included in the switching model by treatment arm and switch status to assess the plausibility of the positivity assumption.*

IPCW3. State whether unstabilised or stabilised weights were used.

IPCW4. Detail the statistical procedure used to calculate weights (e.g. pooled logistic regression, Cox model).

IPCW5. State the portion of data used in the switch model including time-varying predictors (e.g. post-progression data only).

IPCW6. Describe the extent of, and the method used to address, missing data on predictors in the switch model(s).

IPCW7. Present parameter estimates and associated measures of precision from the switch model(s).

IPCW8. Summarise the distribution of weights and state whether values were truncated.

IPCW9. Detail the final outcomes model (i.e. the model fitted to the weighted dataset to estimate the treatment effect, or to model survival), including the estimation method (e.g. robust variance estimation) and the baseline variables adjusted for.

IPCW10. Report on sensitivity analyses showing the robustness of treatment effect estimates and survival extrapolations to violations of key assumptions. Estimated treatment effects and survival extrapolations, AIC, BIC and maximum and minimum switch weights should be compared for versions of the models with different numbers of cubic splines, functional forms, and alternative definitions of categorical variables derived from continuous variables. If weights were truncated, there must be some sensitivity analysis around the truncation percentiles, including a comparison with no truncation.*

RPSFTM

RPSFTM1. Provide a statement around the plausibility of the common treatment effect assumption.*

RPSFTM2. State and justify the structural model assumed (e.g. as treated, ever treated).

RPSFTM3. State the metric used for g-estimation (e.g. log-rank test), including baseline variables for adjustment where applicable.

RPSFTM4. State the grid-search or interval bisection algorithm used for g-estimation.

RPSFTM5. Plot g-estimation results to show that the estimation process has worked well.

RPSFTM6. Present the estimated time ratio (or acceleration factor) and its confidence interval.

RPSFTM7. Compare counterfactual survival times between randomised groups in a Kaplan-Meier plot.

RPSFTM8. Detail the model fitted to the adjusted dataset, including the method used to calculate confidence intervals around the estimated treatment effect and/or survival extrapolation (e.g. retain ITT p-value, bootstrapping) and baseline variables adjusted for.

RPSFTM9. Present results both with and without re-censoring applied.

RPSFTM10. Report on sensitivity analyses showing the robustness of treatment effect estimates and survival extrapolations to violations of key assumptions. Sensitivity analysis around the common treatment effect assumption should be included.

IPE

IPE1. Provide a statement around the plausibility of the common treatment effect assumption.*

IPE2. State and justify the structural model assumed (e.g. as treated, ever treated).*

IPE3. State the parametric model used and consider its appropriateness.*

IPE4. Present the estimated time ratio (or acceleration factor) and its confidence interval.*

IPE5. Compare counterfactual survival times between randomised groups in a Kaplan-Meier plot.*

IPE6. Detail the model fitted to the adjusted dataset, including the method used to calculate confidence intervals around the estimated treatment effect and/or survival extrapolation (e.g. retain ITT p-value, bootstrapping) and baseline variables adjusted for.*

IPE7. Present results both with and without re-censoring applied.*

IPE8. Report on sensitivity analyses showing the robustness of treatment effect estimates and survival extrapolations to violations of key assumptions. Sensitivity analysis around the common treatment effect assumption should be included.*

Two-stage estimation (simple version)

TSEs1. State the disease-related secondary baseline used.*

TSEs2. State the parametric model used and consider its appropriateness.*

TSEs3. Provide a statement around the plausibility of the no unmeasured confounding assumption. This can include a relevant DAG for selection of variables.*

TSEs4. Provide a statement around the plausibility of the assumption of no time-dependent confounding between secondary baseline and time of switch. This should include a graph illustrating the time from secondary baseline to switch across all switching patients.*

TSEs5. Describe the extent of, and the method used to address, missing data on predictors used in the analysis.*

TSEs6. Present the estimated time ratio (or acceleration factor) and its confidence interval.*

TSEs7. Detail the model fitted to the adjusted dataset, including the method used to calculate confidence intervals around the estimated treatment effect and/or survival extrapolation (e.g. bootstrapping) and baseline variables adjusted for.*

TSEs8. Present results with and without re-censoring.*

TSEs9. Report on sensitivity analyses showing the robustness of treatment effect estimates and survival extrapolations to violations of key assumptions.

Estimated treatment effects, survival extrapolations, AIC and BIC should be compared for versions of the models with different parametric distributions and covariates included.*

Two-stage estimation (g-estimation version)

TSEg1. State the disease-related secondary baseline (if used).*

TSEg2. Provide a statement around the plausibility of the no unmeasured confounding assumption. This can include a relevant DAG for selection of variables.*

TSEg3. Provide a graph illustrating the time from secondary baseline to switch across all switching patients.*

TSEg4. Present the switching model and counterfactual survival model used in the g-estimation process.*

TSEg5. State the portion of data used in the switch model including time-varying predictors (e.g. post-progression data only).*

TSEg6. Describe the extent of, and the method used to address, missing data on predictors in the switch model(s).*

TSEg7. State the g-test used (e.g. Wald test, likelihood ratio test, and whether a sandwich variance was used with or without clustering).*

TSEg8. State how the potential survival outcome is entered into the switching model (e.g. only using the event indicator, the time-to-event or censoring, or a combination of these).*

TSEg9. State the grid-search or interval bisection algorithm used for g-estimation.

TSEg10. Plot g-estimation results to show that the estimation process has worked well.*

TSEg11. Present the estimated time ratio (or acceleration factor) and its confidence interval.*

TSEg12. Detail the model fitted to the adjusted dataset, including the method used to calculate confidence intervals around the estimated treatment effect and/or survival extrapolation (e.g. bootstrapping) and baseline variables adjusted for.*

TSEg13. Present results with and without re-censoring.*

TSEg14. Report on sensitivity analyses showing the robustness of treatment effect estimates and survival extrapolations to violations of key assumptions. Estimated treatment effects and survival extrapolations should be compared for versions of the models with different functional forms and covariates.*

Other Treatment Switching Adjustment Methods

OTS1. Provide a statement around the plausibility of the key assumptions.*

OTS2. For methods that require the no unmeasured confounding assumption, provide a statement around the plausibility of the assumption together with a description of the covariate selection approach used – this can include a relevant DAG.*

OTS3. Describe the extent of, and the method used to address, missing data on predictors used in the analysis.*

OTS4. Detail models and any estimation processes used, including method of obtaining confidence intervals around the estimated treatment effects and survival extrapolations.*

OTS5. Report details of model specifications and outputs at each stage.*

OTS6. Provide evidence of model performance, with the approach appropriate to the type of analysis used (e.g. if g-estimation is used, plot g-estimation results; if weighting is used, present parameter estimates and associated measures of precision from the weighting model(s) and summarise the distribution of the estimated weights).*

OTS7. Provide a visual comparison of observed and adjusted survival times.*

OTS8. Report on sensitivity analyses showing the robustness of treatment effect estimates and survival extrapolations to violations of key assumptions.*

OTS9. If external data are used to adjust for treatment switching recommendations OTS1-OTS8 should be followed, as should relevant guideline documents related to the use of matched controls and analysing observational data.*

We expect that improved reporting of treatment switching analyses will facilitate improved review of these analyses by external assessment groups and NICE

appraisal committees. However, to further encourage this, we make the following recommendation:

Recommendation 9: *External assessment groups and appraisal committees should routinely consider the issue of treatment switching.* This should include a consideration of (i) treatment switching that occurred in the pivotal trials and whether these are representative of treatment pathways available in standard clinical practice; (ii) whether the observed treatment switches are likely to impact upon outcomes important for the clinical and cost-effectiveness analysis; (iii) whether adjustments have been made for any types of treatment switching for which adjustment is deemed appropriate; (iv) the case-specific validity of the methods used for any adjustments made. If sufficient information is not provided in the evidence submission, further details should be requested.

We also re-iterate that the method selection advice provided by TSD 16 remains relevant.

Recommendation 10: *Refer to TSD 16[1] as well as this TSD to assist with the identification of adjustment methods that may be appropriate on a case-by-case basis.* Occasionally it may be possible to rule out specific methods a priori, but in general we recommend that multiple treatment switching adjustment methods should be applied to test the sensitivity of the results to the model assumptions. Thorough justification should be provided if any of these methods cannot be applied. It is important that the data requirements for the application of methods are established at the planning stage of the trial, to ensure that necessary data are collected during the trial to allow adjustment methods to be successfully applied.

7. SUMMARY

This TSD provides an update to TSD 16. In this update, we emphasise the importance of including planning for treatment switching adjustment analyses at the outset of any trial, to ensure that appropriate and sufficient data are collected during the trial to perform appropriate adjustment analyses. This is crucial because missing data is a common problem in RCTs, especially in variables measured over time, yet many of the adjustment methods detailed in this document rely crucially on covariate information being available.

We aim to encourage clarity and consistency in the presentation of results and information supporting the analyses, by providing a detailed list of items that should be included alongside treatment switching adjustment analyses in submissions to NICE. We note that the consideration of hypothetical estimands in the ICH E9 R1 Addendum [6] means that adjustment analyses may become more common in submissions made to regulatory agencies and in trial planning, which should further encourage consistency and transparency in the conduct and reporting of these analyses.

TSD 16 focused on treatment switching defined as participants randomised to the control arm of an RCT crossing over onto the experimental treatment. In this new TSD we extend this to consider switching from the experimental arm onto the control treatment, and situations where participants in either trial arm switch onto any other subsequent therapies. It is appropriate to adjust for switches onto subsequent therapies that are not available in standard clinical practice within the jurisdiction for whom the analysis is undertaken. Hence, in submissions to NICE, it is relevant to adjust for switches to treatments that are not part of standard clinical practice in England and Wales at the time of the NICE appraisal. Our review of recent submissions to NICE has indicated that it is common for trial participants to switch onto ‘other’ subsequent treatments, and that this type of switching and switches from the control group onto the experimental treatment can occur in the same trial. This situation may require adjustment for two (or more) treatment effects associated with different types of switching. IPCW and TSE methods are capable of adjusting for several different types of switching within one trial without amendment to the methods,

and can be used to make adjustments for switches in both arms of a trial. The standard one-parameter RPSFTM and IPE, as described in TSD 16, are only capable of adjusting for switches to a non-study treatment under the assumption that the non-study treatment has the same treatment effect as the experimental treatment investigated in the trial.

Key methodological developments that have occurred since publication of TSD 16 include;

a. a more complex version of the TSE method, that can cope with time-dependent confounding.
b. extended versions of the RPSFTM that attempt to estimate more than one treatment effect.
c. enhanced capability to test the sensitivity of RPSFTM and IPE adjustment results to the common treatment effect assumption.
d. new research on the impact of re-censoring.
e. new work on covariate selection, relevant for IPCW and TSE methods.

We regard (a), (c), (d) and (e) to be of particular importance: when the TSE method is used and there is concern that the analysis is prone to time-dependent confounding, the more complex version that uses g-estimation should be applied; when the RPSFTM or IPE methods are used, sensitivity analysis should always be conducted around the common treatment effect assumption; RPSFTM and TSE analyses should always be conducted with and without re-censoring; and details on covariate selection should always be provided when using IPCW or TSE methods. We would not rule out the use of new methods, such as those referred to in (b), but we remain concerned about the ability of multi-parameter RPSFTM models to successfully estimate more than one treatment effect. We consider sensitivity analysis around the common treatment effect assumption to represent a good substitute for this.

7.1 AREAS FOR FURTHER RESEARCH

The issue of treatment switching continues to be an important factor in a substantial proportion of NICE appraisals. Research into treatment switching adjustment methods is ongoing, and new approaches and extensions to methods continue to be developed.

Recent research has discussed the development of an evidence base for new biostatistical methods, including the use of neutral comparison studies.[63] This would be highly valuable in the area of treatment switching adjustment methods. Several of the new methods discussed in this document have not been compared to existing methods in simulation studies or applied analyses, and previously published simulation studies have not considered scenarios that include multiple switches. Further research in these areas could provide further insight on the performance and reliability of adjustment methods, and could be used to update the recommendations contained in this document.

Additional areas of research mentioned in this TSD include:

- For trial planning, research on calculating sample sizes for trials where treatment switching is expected.
- Research on optimal model specifications for IPCW and TSE methods when adjustments are required for multiple types of switching (for example, whether treatments for which adjustments are required should be grouped together, or adjusted for separately)

Finally, given that researchers involved in the health technology assessment process regularly do not have access to patient-level data from clinical trials affected by switching, it would be valuable to further research approaches for adjusting for treatment switching in situations where only summary data are available. Related to this, further research on methods that attempt to adjust for treatment switching using external data would also be valuable, especially using EHR data sources.

REFERENCES

1. Latimer NR, A.K., *NICE DSU Technical Support Document 16: Adjusting Survival Time Estimates in the Presence of Treatment Switching.* . 2014, National Institute for Health and Care Excellence (NICE).
2. Robins, J.M. and A.A. Tsiatis, *Correcting for non-compliance in randomized trials using rank preserving structural failure time models.* Communications in statistics-Theory and Methods, 1991. **20**(8): p. 2609-2631.
3. Robins, J.M. and D.M. Finkelstein, *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): p. 779-788.
4. Branson, M. and J. Whitehead, *Estimating a treatment effect in survival studies in which patients switch treatment.* Statistics in medicine, 2002. **21**(17): p. 2449-2463.
5. Latimer, N.R., et al., *Adjusting for treatment switching in randomised controlled trials–a simulation study and a simplified two-stage method.* Statistical methods in medical research, 2017. **26**(2): p. 724-751.
6. International Council for Harmonisation of Technical Requirements for Human Use (ICH), *E9 (R1) addendum on estimands and sensitivity analysis in clinical trials to the guideline on statistical principles for clinical trials.* 2019.
7. Xu, Y., et al., *Teasing Out the Overall Survival Benefit With Adjustment for Treatment Switching to Multiple Treatments.* Statistics in Biopharmaceutical Research, 2021: p. 1-10.
8. Li, L., S. Tang, and L. Jiang, *On an enhanced rank-preserving structural failure time model to handle treatment switch, crossover, and dropout.* Statistics in Medicine, 2017. **36**(10): p. 1532-1547.
9. White, I.R., S. Walker, and A. Babiker, *strbee: Randomization-based efficacy estimator.* The Stata Journal, 2002. **2**(2): p. 140-150.
10. Latimer, N., et al., *Improved two-stage estimation to adjust for treatment switching in randomised trials: g-estimation to address time-dependent confounding.* Statistical methods in medical research, 2020. **29**(10): p. 2900-2918.
11. Cole, S.R. and M.A. Hernán, *Constructing inverse probability weights for marginal structural models.* American journal of epidemiology, 2008. **168**(6): p. 656-664.
12. Sullivan, T.R., et al., *Adjusting for treatment switching in oncology trials: a systematic review and recommendations for reporting.* Value in Health, 2020. **23**(3): p. 388-396.
13. Monnickendam, G., et al., *Measuring Survival Benefit in Health Technology Assessment in the Presence of Nonproportional Hazards.* Value in Health, 2019. **22**(4): p. 431-438.
14. NICE, *NICE health technology evaluations: the manual*. 2022.
15. Manitz, J., et al., *Estimands for overall survival in clinical trials with treatment switching in oncology.* Pharmaceutical Statistics, 2022. **21**(1): p. 150-162.
16. Clark, T.P., et al., *Estimands: bringing clarity and focus to research questions in clinical trials.* BMJ open, 2022. **12**(1): p. e052953.
17. Henshall, C., et al., *Treatment switching in cancer trials: issues and proposals.* International journal of technology assessment in health care, 2016. **32**(3): p. 167-174.

18. Latimer, N.R., et al., *Assessing methods for dealing with treatment switching in clinical trials: a follow-up simulation study.* Statistical methods in medical research, 2018. **27**(3): p. 765-784.
19. Tennant, P.W., et al., *Use of directed acyclic graphs (DAGs) to identify confounders in applied health research: review and recommendations.* International journal of epidemiology, 2021. **50**(2): p. 620-632.
20. Shinozaki, T. and E. Suzuki, *Understanding marginal structural models for time-varying exposures: pitfalls and tips.* Journal of epidemiology, 2020: p. JE20200226.
21. Rutherford, M.J., et al., *NICE DSU Technical Support Document 21. Flexible Methods for Survival Analysis*. 2020, NICE Decision Support Unit.
22. White, I.R., *Uses and limitations of randomization-based efficacy estimators.* Statistical methods in medical research, 2005. **14**(4): p. 327-347.
23. White, I.R., et al., *Randomization-based methods for correcting for treatment changes: examples from the Concorde trial.* Statistics in medicine, 1999. **18**(19): p. 2617-2634.
24. Allison, A., I.R. White, and S. Bond, *Rpsftm: an R package for rank preserving structural failure time models.* The R journal, 2017. **9**(2): p. 342.
25. Latimer, N., et al., *Causal inference for long-term survival in randomised trials with treatment switching: Should re-censoring be applied when estimating counterfactual survival times?* Statistical methods in medical research, 2019. **28**(8): p. 2475-2493.
26. White, I.R., *Estimating treatment effects in randomized trials with treatment switching.* Statistics in medicine, 2006. **25**(9): p. 1619-1622.
27. Chen, Q., et al., *Estimation of treatment effects and model diagnostics with two-way time-varying treatment switching: an application to a head and neck study.* Lifetime data analysis, 2020. **26**(4): p. 685-707.
28. Bowden, J., et al., *Gaining power and precision by using model–based weights in the analysis of late stage cancer trials with substantial treatment switching.* Statistics in medicine, 2016. **35**(9): p. 1423-1440.
29. Jiménez, J.L., et al., *A modified weighted log-rank test for confirmatory trials with a high proportion of treatment switching.* Plos one, 2021. **16**(11): p. e0259178.
30. Ristl, R., et al., *Delayed treatment effects, treatment switching and heterogeneous patient populations: How to design and analyze RCTs in oncology.* Pharmaceutical statistics, 2021. **20**(1): p. 129-145.
31. Robins, J.M. and S. Greenland, *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): p. 737-749.
32. Hernan, M. and J. Robins, *Causal Inference: What if. Boca Raton: Chapman & Hill/CRC.* 2020.
33. Latimer, N., K. Abrams, and U. Siebert, *Two-stage estimation to adjust for treatment switching in randomised trials: a simulation study investigating the use of inverse probability weighting instead of re-censoring.* BMC medical research methodology, 2019. **19**(1): p. 1-19.
34. Latimer, N.R., et al., *Assessing methods for dealing with treatment switching in clinical trials: A follow-up simulation study.* Stat Methods Med Res, 2018. **27**(3): p. 765-784.

35. Bai, X., et al., *Adaptive truncated weighting for improving marginal structural model estimation of treatment effects informally censored by subsequent therapy.* Pharmaceutical Statistics, 2015. **14**(6): p. 448-454.
36. Latimer, N., et al., *Improved two-stage estimation to adjust for treatment switching in randomised trials: g-estimation to address time-dependent confounding.* Statistical methods in medical research, 2020. **29**(10): p. 2900-2918.
37. Ouwens, M., et al., *Assessing the Influence of Subsequent Immunotherapy on Overall Survival in Patients with Unresectable Stage III Non–Small Cell Lung Cancer from the PACIFIC Study.* Current Therapeutic Research, 2021. **95**: p. 100640.
38. Luo, X., et al., *A proposed approach for analyzing post-study therapy effect in survival analysis.* Journal of Biopharmaceutical Statistics, 2016. **26**(4): p. 790-800.
39. Zeng, D., et al., *Estimating treatment effects with treatment switching via semicompeting risks models: an application to a colorectal cancer study.* Biometrika, 2012. **99**(1): p. 167-184.
40. Zhang, F., et al., *Assessing importance of biomarkers: A Bayesian joint modelling approach of longitudinal and survival data with semi-competing risks.* Statistical Modelling, 2021. **21**(1-2): p. 30-55.
41. Huang, C.H., et al., *Semiparametric copula-based analysis for treatment effects in the presence of treatment switching.* Statistics in Medicine, 2020. **39**(22): p. 2936-2948.
42. Kuehne, F., et al., *Estimating Treatment-Switching Bias in a Randomized Clinical Trial of Ovarian Cancer Treatment: Combining Causal Inference with Decision-Analytic Modeling.* Medical decision making : an international journal of the Society for Medical Decision Making, 2021. **NA**(NA): p. 272989X211026288-NA.
43. Boucher, R.A., KR; Crowther, MJ; Lambert, PC; Wailoo, AJ; Latimer, NR, *PRM201: Adjusting for treatment switching in clinical trials when only summary data are available - an evaluation of potential methods.* Value in Health, 2013. **16**: p. A610-A611.
44. Boucher, R.A., KR; Lambert, P., *Simulating Individual Patient Level Data Using An Illness-Death Modelling Framework In Order to Adjust for Treatment Switching When Only Summary Data Are Available.* Value in Health, 2015. **18**(7): p. PA701.
45. NICE, *TA 171: Lenalidomide for the treatment of multiple myeloma in people who have received at least 2 prior therapies.* 2009.
46. Ishak, K.J., et al., *Adjusting for Patient Crossover in Clinical Trials Using External Data: A Case Study of Lenalidomide for Advanced Multiple Myeloma.* Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research, 2011. **14**(5): p. 672-678.
47. Faria, R.H.A., Monica; Manca, Andrea; Wailoo, Allan J, *NICE DSU Technical Support Document 17: The use of observational data to inform estimates of treatment effectiveness in technology appraisal: Methods for comparative individual patient data. Report by the Decision Support Unit.* 2015.
48. Phillippo, D.A., A.E.; Dias, Sofia; Palmer, Stephen; Abrams, Keith R; Welton, Nicky J, *NICE DSU Technical Support Document 18: Methods for Population-Adjusted Indirect Comparisons in Submissions to NICE. Report by the Decision Support Unit.* 2016.

49. NICE, *NICE real-world evidence framework. Corporate document ECD9.* 2022.
50. NICE, *TA784: Niraparib for maintenance treatment of relapsed, platinum-sensitive ovarian, fallopian tube and peritoneal cancer*. March 2022.
51. NICE, *TA741: Apalutamide with androgen deprivation therapy for treating hormone-sensitive metastatic prostate cancer*. October 2021.
52. NICE, *TA740: Apalutamide with androgen deprivation therapy for treating high-risk hormone-relapsed non-metastatic prostate cancer*. October 2021.
53. NICE, *TA709: Pembrolizumab for untreated metastatic colorectal cancer with high microsatellite instability or mismatch repair deficiency*. June 2021.
54. NICE, *TA705: Atezolizumab monotherapy for untreated advanced non-small-cell lung cancer*. June 2021.
55. NICE, *TA660: Darolutamide with androgen deprivation therapy for treating hormone-relapsed non-metastatic prostate cancer*. November 2020.
56. NICE, *TA653: Osimertinib for treating EGFR T790M mutation-positive advanced non-small-cell lung cancer*. October 2020.
57. NICE, *TA742: Selpercatinib for treating advanced thyroid cancer with RET alterations*. November 2021.
58. NICE, *TA670: Brigatinib for ALK-positive advanced non-small-cell lung cancer that has not been previously treated with an ALK inhibitor*. January 2021.
59. NICE, *TA643: Entrectinib for treating ROS1-positive advanced non-small-cell lung cancer*. August 2020.
60. NICE, *TA528: Niraparib for maintenance treatment of relapsed, platinum-sensitive ovarian, fallopian tube and peritoneal cancer*. July 2018.
61. BMJ TAG, *Niraparib for maintenance treatment of platinum-sensitive ovarian cancer after second response to chemotherapy (CDF review of TA528). ERG response to company ACD comments* 2021.
62. Diels, J., et al., *PCN248 Estimating the unbiased OS-benefit for apalutamide (APA) versus androgen deprivation therapy (ADT) in patients with non-metastatic castration-resistant prostate cancer (NMCRPC) from SPARTAN, adjusting for subsequent therapy using randomized data external to the trial.* Value in Health, 2019. **22**: p. S103.
63. Heinze, G., et al., *Phases of methodological research in biostatistics—Building the evidence base for new methods.* Biometrical Journal. **n/a**(n/a): p. 2200222.
64. Schlosser, R.W., et al., *Use of information-seeking strategies for developing systematic reviews and engaging in evidence-based practice: the application of traditional and comprehensive Pearl Growing. A review.* International Journal of Language & Communication Disorders, 2006. **41**(5): p. 567-582.
65. Alshreef, A., et al., *Statistical methods for adjusting estimates of treatment effectiveness for patient nonadherence in the context of Time-to-Event outcomes and health technology assessment: a systematic review of methodological papers.* Medical Decision Making, 2019. **39**(8): p. 910-925.
66. Haddaway, N., M. Grainger, and C. Gray, *citationchaser: An R package and Shiny app for forward and backward citations chasing in academic searching*. 2021.

APPENDIX

Appendix A: Methods for the systematic review of the literature to inform Section 4 of the TSD

Review Question: The review aimed to identify methods that have been used to adjust for treatment switching since the original TSD on treatment switching (TSD 16) was published in 2014.

Search strategy: The search took two approaches. Approach 1 followed the traditional style of systematic review and approach 2 used a Comprehensive Pearl Growing technique.[64, 65]

1. Searches were performed in PubMed, Scopus and Web of Science on 22/12/2021, using the following search terms - ("treatment switching" OR "treatment crossover") AND "trial" AND "adjust*" NOT "crossover trial".
2. The base set of treatment switching methods papers were identified from the bibliography of TSD 16. A search was performed to identify all papers that cite these papers by uploading DOIs for each of these papers into CitationChaser, and using the forward citation chasing search function. [66]

The results from these searches were combined and duplicates were dropped. See Figure A1.

Inclusion and exclusion criteria: We excluded papers that were published prior to 2014, before the publication of TSD 16. Papers that were unpublished, applied methods as described in TSD 16 without extension, or methods applied to observational data were excluded. We included papers published in peer reviewed journals, that described new methods that adjust survival for treatment switching in clinical trials or extended the methods described in TSD 16. Eligibility was initially assessed based on abstract. Full texts were reviewed for those with eligible abstracts.

Figure A1: PRISMA flow diagram for literature search to identify switching adjustment methods