NICE DSU TECHNICAL SUPPORT DOCUMENT 19:

PARTITIONED SURVIVAL ANALYSIS FOR DECISION MODELLING IN HEALTH CARE: A CRITICAL REVIEW

REPORT BY THE DECISION SUPPORT UNIT

2 June 2017

Beth Woods[1], Eleftherios Sideris[1], Stephen Palmer[1], Nick Latimer[2], Marta Soares[1]

[1]Centre for Health Economics, University of York, York, UK
[2] School of Health and Related Research, University of Sheffield, 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 www.nicedsu.org.uk
Twitter @NICE_DSU

About the Decision Support Unit

The Decision Support Unit (DSU) is a collaboration between the Universities of Sheffield, York and Leicester. We also have members at the University of Bristol, London School of Hygiene and Tropical Medicine and Brunel University. The DSU is commissioned by the National Institute for Health and Care Excellence (NICE) to provide a research and training resource to support the Institute's Technology Appraisal Programme. Please see our website for further information www.nicedsu.org.uk

About the Technical Support Document series

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

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

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

Dr Allan Wailoo

Director of DSU and TSD series editor.

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

ACKNOWLEDGEMENTS

The DSU thanks Adrian Bagust, Jaclyn Beca, Ewen Cummins, Martin Hoyle, Chris Jackson, Paul Tappenden, Claire Williams and the team at NICE for reviewing this document. The production of this document was funded by the National Institute for Health and Care Excellence (NICE) through its Decision Support Unit. The views, and any errors or omissions, expressed in this document are of the authors only. NICE may take account of part or all of this document if it considers it appropriate, but it is not bound to do so.

This report should be referenced as follows:

Woods B, Sideris E, Palmer S, Latimer N, Soares M. NICE DSU Technical Support Document 19. Partitioned Survival Analysis for Decision Modelling in Health Care: A Critical Review. 2017 [Available from http://www.nicedsu.org.uk]

EXECUTIVE SUMMARY

Cost-effectiveness analyses informing NICE appraisals use a range of modelling approaches such as decision trees, Markov models and individual sampling models. Most of these have been subject to detailed discussion within the economic evaluation literature. This has not been the case for one approach, partitioned survival analysis, which has been used extensively in the NICE Technology Appraisal (TA) Programme and is now the most commonly used decision modelling approach for appraisals of interventions for advanced or metastatic cancers.

The objective of this Technical Support Document (TSD) is to describe and critique partitioned survival analysis when used as a decision modelling tool in order to assist different stakeholders in determining its appropriateness as a basis for informing policy decisions. This TSD provides:

- A description of the partitioned survival analysis approach and how it differs from more conventional state transition models in terms of structural assumptions and data requirements (Section 2).
- A review of the use of partitioned survival analysis in recent NICE TAs of cancer treatments (Section 3).
- A critique of the approach focusing on the implications of the structural assumptions made for extrapolation and quantification of uncertainties (Section 4).
- The relative merits of partitioned survival analysis and state transition modelling approaches (Section 4).
- Recommendations for the selection of a modelling approach, documentation of the selected approach, representation of uncertainties relating to extrapolation, and for further methods research (Section 6).

The focus of this TSD is the application of partitioned survival analysis in appraisals of cancer treatments as this is where the approach has been used most frequently to date. Nonetheless, the TSD has broader relevance for the application of partitioned survival analysis as a decision modelling approach in other clinical settings.

CONTENTS

Abbreviations

AG	Assessment Group
AUC	Area under the curve
ERG	Evidence Review Group
HRQoL	Health-related quality of life
IPCW	Inverse probability of censoring weights
IPD	Individual patient data
IPE	Iterative parameter estimation
MTA	Multiple Technology Appraisal
NICE	National Institute for Health and Care Excellence
OS	Overall survival
PartSA	Partitioned survival analysis
PFS	Progression-free survival
PPS	Post-progression survival
PSA	Probabilistic sensitivity analysis
PSM	Progressed state membership
QALYs	Quality-adjusted life years
Q-TWIST	Quality-adjusted time without symptoms or toxicity
RPSFT	Rank preserving structural failure time
STA	Single Technology Appraisal
TA	Technology Appraisal
TSD	Technical Support Document
TTD	Time to treatment discontinuation

1. INTRODUCTION

Decision modelling has an established role in informing clinical decision making and policy decisions relating to cost-effectiveness. When using decision models to generate estimates of the costs and effects of interventions, a range of alternative decision modelling approaches can be used, such as decision trees, Markov models and individual sampling models. Existing literature considers the relative merits of these approaches (e.g. Sonnenberg *et al.* 1993, [1] Barton *et al*. 2004 [2] and Brennan *et al*. 2006 [3]). Economic evaluations have also used a method not considered in this literature, a method referred to as partitioned survival analysis (PartSA) or area under the curve (AUC) modelling. This approach has been used in appraisals of interventions that are expected to prolong life expectancy and impact upon quality of life, mainly in advanced or metastatic cancers (e.g. Hoyle *et al*. 2013;[4] Hornberger *et al*. 2010;[5] Delea *et al*. 2014[6]) although there are examples of its application in other areas including haematological cancers with relatively favourable prognosis.[7] The strengths and limitations of this approach to decision modelling have not been formally considered.

The objective of this Technical Support Document (TSD) is to provide a clear description and critique of PartSA as a decision modelling approach to inform policy decisions relating to cost-effectiveness. A set of recommendations is proposed to aid those developing and reviewing decision models (companies, Assessment Groups, AGs and Evidence Review Groups, ERGs) and making decisions based on their outputs (Appraisal Committees) to better understand the assumptions underlying the approach, and its strengths and weaknesses when compared with alternative decision modelling approaches. Areas of future methods research that would further inform the selection and implementation of appropriate modelling approaches are also outlined.

This TSD is set out as follows. Section 2 describes the PartSA approach and how it differs from more conventional state transition models. Section 3 documents how the method has been used in NICE appraisals, focusing on recent oncology appraisals where the method has been most frequently applied. Section 4 provides a critique of PartSA and discusses alternative approaches, specifically state transition models. Section 5 provides a summary of the conclusions of the TSD and Section 6 provides specific recommendations.

2. PARTITIONED SURVIVAL AS A DECISION MODELLING APPROACH

2.1. OVERVIEW OF DECISION MODELLING

Decision modelling provides a quantitative framework for synthesising available evidence, and generating estimates of clinical and cost-effectiveness that are relevant to decision makers.[8] Three important roles of decision modelling in the NICE appraisal process are to allow all relevant interventions to be simultaneously compared, the full range of relevant information to be reflected in a single coherent analysis, and observed information to be used to predict outcomes and costs over an appropriate model time horizon via extrapolation. Decision models also play an important role in allowing heterogeneity and uncertainty to be characterised and reflected in the decision making process.

Decision models aim to describe key biological or clinical processes, and the way in which interventions affect these processes. Many conditions can be described in terms of a series of distinct events (states) that individuals experience and move between. In these situations, state transition models are often used as they are conceptualised in these terms.[9] In state transition models, the allowed movements between states are referred to as transitions, and the speed at which these transitions occur as transition probabilities or rates. State values (sometimes called "rewards") are used to reflect the costs and health-related quality of life (HRQoL) implications of residing in, or transiting between, each health state. Estimates of expected costs and quality- adjusted life years (QALYs) are derived by assigning state values to the time spent by patients in each health state.

2.2. PARTITIONED SURVIVAL ANALYSIS (PARTSA)

PartSA models are conceptually similar to state transition models in that they are characterised by a series of health states with associated state values. However, they differ in the way that the proportion of patients in each health state at each time point (state membership) is determined. In state transition models, state membership is usually determined using matrices of transition probabilities which describe the probability an

individual will make each transition in a given time period. In the PartSA approach, state membership is determined from a set of non-mutually exclusive survival curves. The PartSA approach uses an overall survival (OS) curve to estimate the proportion of people alive over time directly and may include a statistical extrapolation beyond the time horizon of the original study depending on the requirement to model a lifetime horizon and the maturity of the available data. The area under the (extrapolated) OS curve provides an estimate of mean life expectancy. OS may be further disaggregated or "partitioned" into different health states to allow these health states to have different HRQoL and cost implications. Within PartSA models, there is a survival curve for each health state that describes time from model start (i.e. patient entry in to the model) to transiting to any health state that is further along the sequence. This means that the survival curves do not represent mutually exclusive estimates of state membership. For example, the survival curve for the last alive health state in the sequence - the OS curve - will include all patients who are still alive, including those in earlier health states in the sequence. When there are multiple health states in which patients who are alive can reside, state membership must therefore be derived from the modelled survival curves. State membership is not estimated using transition probabilities that describes transitions from one health state to the next; it is instead derived from the set of non-mutually exclusive survival curves.

The way in which state membership is determined in PartSA models can be illustrated using a model structure commonly applied in economic evaluations of treatments for advanced or metastatic cancer. This model includes three states: progression-free, progressed and dead. Progression implies a worsening or spreading of the cancer. The definition of progression depends on the type of cancer and the criteria used. Patients who are alive are disaggregated according to progression status as progression generally has implications for HRQoL and costs. The PartSA approach requires two survival curves to estimate state membership for this model. The first describes time from model entry to exiting the progression-free state via progression or death (a composite outcome) and the second time from model entry to death. These are the progression-free survival (PFS) and OS endpoints commonly reported in cancer trials. The first survival curve (PFS) directly provides the proportion of patients remaining in the first health state – progression-free – over time. State membership for the dead state is simply 1 minus the OS curve at each time point. For the progressed health state, state membership is derived as the difference between the OS and the PFS curve at each time point, as this provides the proportion of

patients who are alive but not progression-free. This process of deriving the progressed state membership, *PSM(t)* is illustrated in Figure 1. Differences between interventions are typically modelled by using different PFS and OS curves for each treatment.

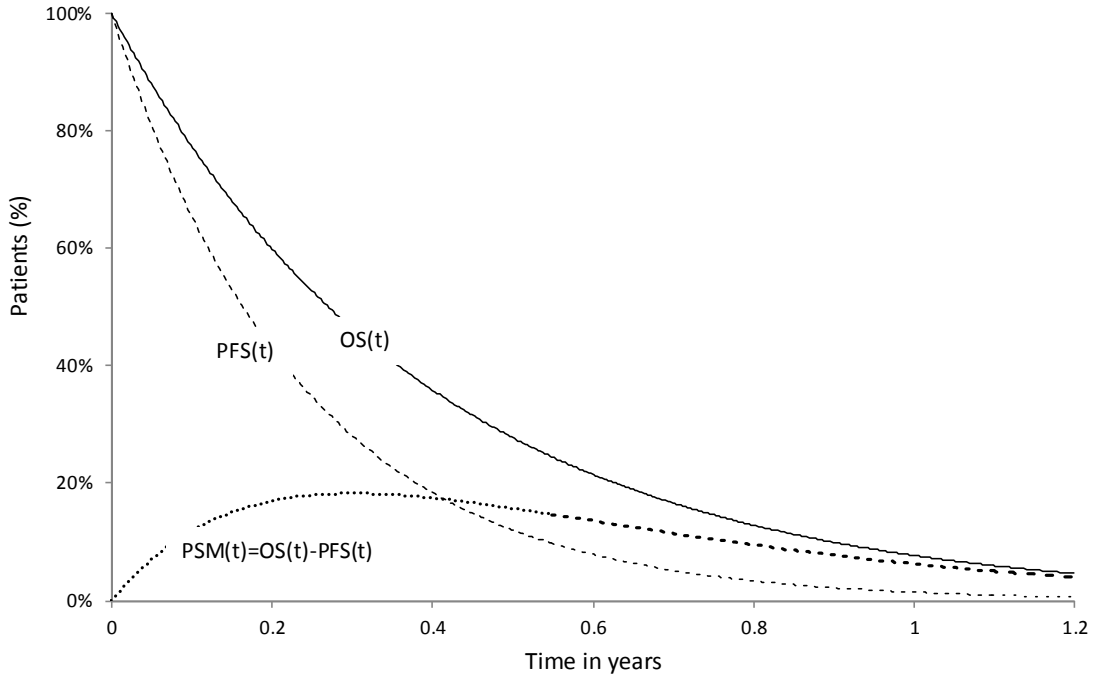


Figure 1: Determining state membership in partitioned survival analysis models, an example of a three-state cancer model [adapted from the Pazopanib company submission to NICE].[10]
PSM(t) denotes progressed state membership (PSM) as a function of time (t).

Although many applications of the PartSA approach use a three-state structure as described in Figure 1, additional states can be included within PartSA models. For example, if patients who progress following initial treatment receive further treatment, four health states could be used: progression-free on initial treatment, progression-free on further treatment, progressed following further treatment and dead.

More generally, we can consider an *N* state PartSA model where the order of states from 1 to *N* (dead) is determined by the order in which patients can move through the states and patients cannot move to a lower indexed health state. For example, for the three state example described above, patients cannot move from being progressed to being progression-free. To specify the PartSA model *n=1,…,N-1* survival curves are required. For health state *n* the corresponding survival function ($S(t)_n$, *also referred to as the cumulative*

survival probability or function) describes time-to-event data from *model entry* to transiting to *any health state that is further along the sequence than state n*. State membership for the first health state is determined directly from the survival function $S(t)_1$. State membership for each other alive health state ($n=2,\ldots,N-1$) is derived using the survival function for that state ($S(t)_n$) and subtracting from this the survival function for the prior health state ($S(t)_{n-1}$) at each time point. The resulting proportion represents the proportion of individuals who have not entered a higher state than *n*, minus the proportion residing in states lower than *n*. State membership for the dead state is estimated as one minus the final ($S(t)_{N-1}$) survival function which is always the OS curve. The process of deriving state membership from each survival curve is summarised in Table 1.

Table 1: Determining state membership in an N health state partitioned survival analysis model

Health state	State membership at time t
$n=1$	$S(t)_n$
$n>1$ and $n \neq N$	$S(t)_n - S(t)_{n-1}$
$n=N$	$1-S(t)_{n-1}$

$S(t)_n$ is a survival function describing the probability that a patient survives in state *n* or a lower indexed state beyond a specific time point (*t*) from model entry.

Estimates of state membership can be represented visually as the observed differences between the survival curves at specific time points. This is shown for a four-state example in Figure 2 where state 4 is death and state membership estimates are shown for the 0.6 year time point using vertical lines.

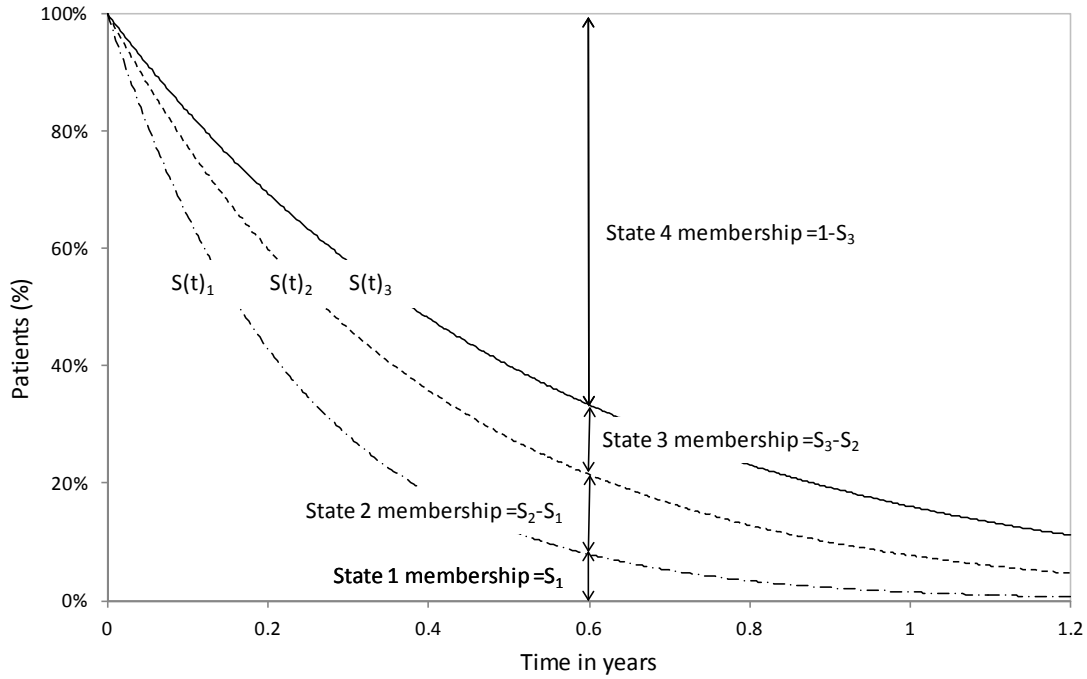


Figure 2: Determining state membership in partitioned survival analysis models, an example of a four state model.

$S(t)_n$ is a survival function describing the probability that a patient survives in state *n* or a lower indexed state beyond a specific time point (*t*) from model entry.

PartSA originated as a method for summarising the overall impact of treatments on HRQoL and survival, in the context of clinical trials. The method was suggested in 1989-1990 by Goldhirsch *et al*. and Glasziou *et al*. [11, 12] who noted that analysing quality-adjusted survival using standard survival analysis methods induced informative censoring. This occurred, as individuals with worse HRQoL were more likely to be censored earlier. The authors developed the PartSA approach to overcome this and provide trial-based estimates of quality-adjusted survival. A specific variation on this approach called the quality-adjusted time without symptoms or toxicity (Q-TWiST) method partitions individuals according to whether they are symptom- and toxicity-free or not. This method has been used in a wide variety of oncology settings to provide a trial-based summary measure of treatment benefit that takes into account the impact of treatment on survival, disease-related symptoms and the adverse effects of treatment.[13]

The survival curves used in the PartSA approach (e.g. the PFS and OS curves) are derived from independent analyses of each time-to-event endpoint. The PartSA approach was developed as a clinical tool for within-trial analysis and its use as a decision modelling approach has required extensions to the original method to allow for extrapolation and incorporation of additional evidence. In the context of NICE Technology Appraisals (TAs), decision models using the PartSA approach are typically based on one or more "pivotal" trials evaluating the appraised technology or technologies. For these trials, individual patient data (IPD) are generally available to companies or can be estimated by AGs or ERGs from reported summary Kaplan-Meier estimates using established methods.[14, 15] For interventions included in the pivotal trial(s), the survivor functions required for application of PartSA can therefore be estimated using a wide range of analytic options. For the within-trial period, non-parametric approaches (e.g. Kaplan-Meier estimates) or parametric models (e.g. Weibull) may be used, whilst for the extrapolation period, parametric models are generally used. Flexible parametric models or mixture models may be used to allow more complex hazard functions to be represented. Different assumptions regarding the treatment effect can be reflected including proportional hazards or a constant acceleration factor, or allowing time-dependent treatment effects by fitting separate survival models to each arm of a trial.[16] Where external evidence on comparative effectiveness is used, it is generally taken from a meta-analysis, indirect comparison or network meta-analysis. This is commonly incorporated within the model by applying a constant hazard ratio to the hazard for the comparator used as the reference comparator in the synthesis work. Synthesis methods which allow for non-proportional hazards are emerging and can be reflected within PartSA models.[17, 18] Where available, external evidence may be used to help inform the long-term extrapolation of survival.

2.3. CONCEPTUAL DIFFERENCES BETWEEN PARTSA AND STATE TRANSITION MODELS

Although there are clear similarities between the PartSA approach and state transition modelling, there are important conceptual differences. Firstly, the approaches differ in the types of disease and clinical processes to which they apply. PartSA models can only be applied to processes in which patients move forward through a set of health states (though patients do not have to pass through every state), whereas state transition models can be used to represent any specified transitions. In the example model of advanced or metastatic

cancer described above (see Figure 1), a PartSA approach is possible as it is generally considered reasonable to assume that progressed patients cannot return to the original progression-free state (though they can be considered progression-free on a subsequent line of therapy, as progression is generally defined relative to measurements taken at the start of a course of treatment).

Even when modelling the same set of allowed movements between states, PartSA and state transition modelling differ fundamentally in their underlying structural assumptions. In the PartSA approach, each endpoint (e.g. PFS, OS) is modelled independently of the other endpoints included within the model, whereas in state transition models clinical events are explicitly related.

Returning to the example in advanced or metastatic cancer, Figure 3 shows the state transition model that corresponds to the PartSA model shown in Figure 1. State membership for state transition models is most commonly derived using cohort simulation in discrete time. Briefly, this involves specifying a set of transition probabilities $p(c)_{ij}$, which describe the probability of transiting from a certain health state *(i)* to another *(j)* within one discrete time period (a model cycle, *c*). In this example, three transition probabilities would need to be estimated: the probability of disease progression $p(c)_{12}$, the probability of death conditional on being in the progression-free state $p(c)_{13}$ and the probability of death conditional on being in the progressed state $p(c)_{23}$. All patients start the model in the progression-free state. State membership at the end of each cycle is estimated by applying the transition probabilities to the state membership at the end of the previous cycle. This is repeated until the model time horizon is reached.

In some situations, (e.g. when transition probabilities depend on time spent by patients in an intermediate health state such as progressed), cohort simulation using standard methods can become cumbersome. When this is the case, patient-level simulation represents an alternative method for estimating state membership. Further guidance on this approach is available in TSD 15.[19]

State transition models evaluated using cohort or patient-level simulation methods are centred around health states and the transitions between these health states. For the purposes of this TSD the conceptual differences between state transition models and

PartSA are therefore the same, regardless of whether state transition models are implemented using cohort- or patient-level simulation, or other methods.

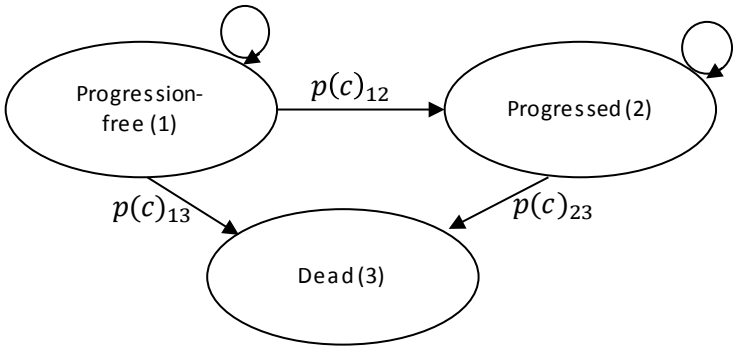


Figure 3: A typical state transition model used in advanced and metastatic cancer.

Arrows indicate allowed transitions including the possibility that individuals will remain in the same state, arrow labels denote the probability pij of transitioning from state i to j in a given cycle, c.

In state transition models, OS depends on all three individual transitions and the rate of death reflects the evolving proportion of patients in the progressed state and the differences in mortality between progression-free and progressed patients. There is therefore a structural link between mortality and earlier progression events. In contrast, mortality in a PartSA model is only determined by time to death data and is not explicitly linked to earlier progression events. This is because in the PartSA approach each modelled endpoint is structurally independent of the other modelled endpoints.

In the context of a within-trial analysis or a case in which data have been fully observed, PartSA and state transition modelling approaches are expected to produce similar results if modelling and fitting have been done appropriately, as relationships between endpoints are reflected within the data. However when data are incomplete and parameters derived from the observed data are used for extrapolation, the approaches are expected to differ. In the PartSA approach, the structural independence between endpoints means that extrapolations for a given endpoint reflect within-trial trends in that endpoint alone, unless external information on hazard rates beyond the trial period is incorporated. If external information is not used, extrapolation of OS depends only upon prior trends in mortality rates, and is not explicitly linked to information on non-fatal events (e.g. progression). Non-fatal events

only impact upon mortality in the extrapolation period to the extent that they influenced the time-trend in mortality in the within-trial period and to the extent that this is reflected by the survival model chosen (which is likely to depend in part on its flexibility in representing changing hazards over time). In a state transition model, extrapolations of OS are influenced by the model structure and the within-trial estimates of each transition probability. In the example shown in Figure 3, OS in the extrapolation period will therefore be influenced by the state membership at the start of the extrapolation period, the rate of progression and how progression status impacts upon mortality. As with PartSA models, external information may be used to inform longer term transition rates in state transition models.

Within PartSA, differences in state membership between treatments are determined by differences in the survival curves between treatments. For the example shown in Figure 1, both PFS and OS endpoints could differ between treatments. Differences in survival outcomes predicted by the model (e.g. PFS, OS) between treatments simply arise from the differences in the survival curves chosen for these independently modelled events for each treatment group.

In contrast, in a state transition model the combined effect of treatment effects on individual transition probabilities and the structural relationship between events determines the treatment effect on PFS and OS. For example, a beneficial treatment effect on OS could be observed due to: (i) a beneficial treatment effect on the probability of progression [$p(c)_{12}$] only (implying a surrogacy relationship between progression and mortality), (ii) a beneficial treatment effect on both $p(c)_{12}$ and the probability of death from the progressed state [$p(c)_{23}$] or (iii) a beneficial treatment effect on $p(c)_{12}$ and a detrimental effect on $p(c)_{23.}$ Treatment effects on earlier transition probabilities may also have indirect effects on subsequent transition probabilities. For example, patients who take longer to progress may have better outcomes with post-progression treatments as they have a longer rest period without anti-cancer therapy. If this is reflected in the model, then patients who take longer to progress will experience a lower probability of death from the progressed state [$p(c)_{23}$]. Any treatment exerting a beneficial effect on the probability of progression [$p(c)_{12}$] will therefore indirectly extend time spent in the progressed state (which we call post-progression survival, *PPS*). Specification of treatment effects for the extrapolation period

in a state transition model therefore requires an estimate of how treatment is expected to impact on each individual transition probability during this period.

Finally, the PartSA and state transition modelling approaches differ in the information required to parameterise the models. For the context shown in Figure 1, the PartSA approach requires estimates of the PFS and OS curves for each treatment, whereas the corresponding state transition model shown in Figure 3 requires time-to-event data on each individual transition probability, and estimates of the effects of each treatment on each transition probability. This is important as cancer trials and other clinical studies typically report only PFS and OS i.e. the data required for the PartSA model. These data cannot be used to derive the time-to-event data required for state transition modelling in a straightforward manner, as PFS describes the combination of progressions and deaths from the progression-free state and OS describes the overall probability of death, which is a function of all three transition probabilities. We return to the challenges of using available data to parameterise alternative model structures in Section 4.2.

3. REVIEW OF NICE CANCER APPRAISALS

As PartSA models were originally proposed as an approach for within-trial analysis, we sought to review applications of the PartSA approach in the context of decision modelling to inform policy decisions relating to cost-effectiveness. The aim of this review was to establish how the PartSA approach is used in this context. Specifically we aimed to establish whether the assumptions made in the PartSA approach were recognised and critiqued, whether the advantages and disadvantages of the approach compared to alternative approaches were recognised and informed the selection of the modelling approach, and how the method was being implemented in terms of assumptions made and data used. The review focuses on cancer appraisals, as a scoping review suggested that this is the indication in which the PartSA method has most commonly been used. We reviewed appraisals using both the PartSA approach and alternative methods to establish the nature of alternative methods used.

3.1. METHODS

We reviewed the 30 most recent published completed NICE TAs of cancer treatments covering the period May 2013-February 2016. This review period was pragmatically selected to identify a sufficient number of appraisals to provide a representative sample of the modelling approaches currently applied to cancer appraisals. A full list of the reviewed TAs is presented in Appendix 1.

Company submissions, assessment reports by academic groups (ERGs or AGs) and the Final Appraisal Determination documents from the NICE website were reviewed. Models were categorised according to the definitions in Box 1. For the PartSA models, we extracted data on the description of and justification for the approach, the survival endpoints modelled and endpoints for which differences between treatments were modelled, the use of data external to the pivotal trial, whether parameter estimates were adjusted to model differences between the trial data and modelling context, whether the probabilistic sensitivity analysis (PSA) accounted for correlation between modelled endpoints, and the main concerns relating to the modelling approach raised by the ERG and the Appraisal Committee. For those appraisals using non-PartSA approaches, we extracted information on the model type, the justification for using the selected approach, the health states modelled, the transitions probabilities assumed to differ by treatment, use of data external to the pivotal trial and the main concerns relating to the modelling approach raised by the ERG and the Appraisal Committee.

Box 1: Definitions of model types.

Note that the Markov and semi-Markov approaches described are Markov chains rather than Markov processes as they are evaluated in discrete time. Note that patient level simulation as described here is assumed to be evaluated using discrete event simulation; an alternative would be to use microsimulation in discrete time though this is generally considered to be computationally inefficient.

1 Partitioned survival analysis, also referred to as Area Under the Curve approach or model.

Individuals in the model reside in one of a series of mutually exclusive and jointly exhaustive health states. State membership is determined fully by a series of independently modelled non-mutually exclusive survival curves. A survival curve must be specified for each alive health state that describes time from *model start* (i.e. patient entry in to the model) to transiting to *any health state that is further along the sequence.*

2 State transition models

Individuals in the model reside in one of a series of mutually exclusive and jointly exhaustive health states. The method of determining state membership is based on specifying which transitions occur and the speed rate at which these occur. There are alternative methods for evaluating state transition models: 2A, 2B and 2C.

2A Markov

State membership is determined using cohort simulation in discrete time and all transition probabilities are constant or depend only on calendar time (i.e. time in model).

2B Semi-Markov

State membership is determined using cohort simulation in discrete time and one or more transition probabilities depend on time in an intermediate health state. Transition probabilities can also depend on calendar time.

2C Patient level simulation

Model is evaluated for a series of individual patients, typically by sampling the time spent in a specific state from corresponding survival curves. Event rates can depend on calendar time or time in state.

3 Decision tree

All patients follow one of a series of mutually exclusive and jointly exhaustive pathways. Each pathway is associated with a probability and pathway values.

3.2. REVIEW FINDINGS

PartSA was used in 22 (73%) of the 30 appraisals reviewed. Only two of the reviewed TAs were Multiple Technology Appraisals (MTAs), while all remaining TAs were Single Technology Appraisals (STAs). The TAs reviewed related predominantly to advanced and metastatic cancers, though some included locally advanced patients. Two of the TAs using non-PartSA approaches related to the adjuvant setting.

The implementation of PartSA within these appraisals is summarised in Table 2 (Appendix 2 provides further details). The remaining eight TAs that used alternative modelling methods are discussed in Section 4.2 where we discuss alternative modelling approaches in more detail.

3.2.1. Description of modelling approach

Despite the frequent use of PartSA, only 12 of the 22 company submissions correctly identified the use of PartSA, while the remaining submissions incorrectly described the approach used as a Markov or semi-Markov model. The ERG described the use of PartSA correctly in 12 of the 22 TAs (though not the same 12 as correctly described by the companies), while in four appraisals the ERG only provided a narrative description of the model and did not specify a modelling approach. In the remaining six appraisals, the method was described incorrectly by the ERG.

3.2.2. Justification for modelling approach

Limited justification was provided for using PartSA. Although all 22 TAs mentioned precedent as a justification, only six companies (and the AG for the MTA) provided any additional reasoning for the use of PartSA. The justifications provided were that this type of modelling approach aligns with the endpoints of the pivotal trials, allows reflection of time-dependency in risks and facilitates the incorporation of external data. In the MTA (TA374), the AG mentioned that PartSA was used due to the routine reporting of PFS and OS data and that ideally they would have implemented an alternative modelling method if PPS data (i.e. time-to-event data describing time from progression to death) were available to populate the model.

3.2.3. Discussion of model assumptions

One of the fundamental assumptions made in PartSA models, that the clinical endpoints are modelled independently, was only acknowledged in four appraisals (TA299, TA344, TA371, TA374) which highlighted only that the approach assumes no correlation between survival endpoints in the PSA.

3.2.4. Survival endpoints modelled

Seventeen of the 22 TAs used two survival endpoints to implement the PartSA model. In the majority of TAs, the modelled survival endpoints were PFS and OS, and these were used to partition individuals into three health states (progression-free, progressed and dead). In two appraisals (TA316, TA377), time to treatment discontinuation or death (TTD) was modelled and served as a proxy for disease progression. The remaining five TAs used three survival endpoints; TTD, PFS and OS. TTD was included to account for the time spent on the initial treatment and patients were partitioned into four states: on-treatment and progression-free, off-treatment and progression-free, progressed and dead.

Additional non-partitioned analyses were used to capture cost and/or HRQoL changes in some of the appraisals, and in particular, to capture the cost and HRQoL implications of subsequent lines of treatment (TA344, TA377), palliative care (TA359) and more advanced phases of the disease (TA299).

3.2.5. Application of treatment effects

In all 22 appraisals that used PartSA, every modelled survival endpoint reflected differences between treatments. In nineteen of these appraisals, the treatment effect was applied for the whole time period modelled, implying that treatment modified the risk of each endpoint for the full model time horizon. The remaining three TAs assumed no treatment effect for specific periods; TA366 and TA321 assumed equal long-term mortality rates between treatments, informed by registry data, while in TA344, OS was extrapolated from trial data and mortality rates were assumed to be the same across treatments following the trial follow-up period. The assumption of equal mortality rates can be applied from a specific time point by applying the same hazards to both trial arms

for all subsequent time points, calculating the resulting cumulative hazard and converting this to a survival estimate using standard formulae.[20]

3.2.6. Data sources

In only five (23%) of the 22 TAs, survival curves from the pivotal trial(s) alone were extrapolated to inform the PartSA. The majority of TAs therefore used some form of external data i.e. data not collected within the pivotal trial. All of the remaining 17 TAs used some form of indirect comparison (naïve or adjusted) to incorporate comparators not included in the pivotal trial. Registry data or pooled long-term survival data from multiple trials were used to inform mortality event rates in the extrapolation period in four of the appraisals (TA321, TA344, TA357, TA366).

3.2.7. Adjustment for treatment switching

In six TAs, the parameter estimates for OS reflected adjustments to attempt to remove the effects of treatment switching, whereby patients who experience disease progression in the control arm of a trial receive treatment with the investigational therapy. The adjustment attempts to remove the potential confounding effect of this treatment switching on the OS estimates for the control therapy. This may be considered appropriate if post-progression administration of the investigational product is not expected in clinical practice, and the methods for treatment switching adjustment are appropriately applied (i.e. when assumptions are considered to be valid).[21]

3.2.8. Other adjustments to parameter estimates

In TA295, an adjustment was made to the parameter estimates, to prevent the extrapolated PFS curve being higher than the OS curve, as this would lead to a logical inconsistency whereby a larger proportion of patients are alive and progression-free than alive. This is possible in PartSA models as PFS and OS are estimated and modelled independently and do not therefore reflect the structural dependence between endpoints. In TA295, the company used a ratio from a published review that translated median PFS benefit into a median OS difference. By applying this published ratio to their PFS data, the company generated an

adjusted OS estimate. The approach was criticised by the ERG and was not accepted by the Committee.

3.2.9. Correlation between endpoints

Correlation between survival curves in PartSA models occurs because the same events contribute to multiple survival curves (e.g. deaths without a prior recorded progression are included in both PFS and OS analyses) and time spent in earlier states contributes to the time to event data for later states (e.g. time spent progression-free contributes to time to death). This correlation was discussed by the company or ERG/AG in only a small number of appraisals (TA285, TA299, TA321, TA344, TA371, TA374). Only two TAs (TA285, TA321) attempted to incorporate any correlation between survival curves in the PSA, all other appraisals modelled the endpoints as being statistically independent. In TA321, a bootstrapping approach was used and it appears that parametric survival models for PFS and OS were repeatedly fitted to bootstrap re-samples of the IPD to generate a correlated distribution of parameters for the PFS and OS models. In TA285, the company adjusted the probabilistic results by assuming that OS is equal to PFS for the probabilistic simulations in which PFS exceeded OS. This ERG noted that this approach was inappropriate.

3.2.10. Concerns raised by ERG/NICE committee

The most commonly raised concerns related to the maturity of the data and the uncertainty around the modelled long-term OS projections. In PartSA models, PPS is calculated as the difference between the OS and PFS curves. The nature of the extrapolated PFS and OS curves for each treatment therefore determines the difference between treatments in PPS. The differences across treatments in the PPS predicted by PartSA models were key points of criticism raised in five appraisals (TA299, TA307, TA309, TA344, TA359). The ERG noted the lack of evidence to support a post-progression treatment benefit in these appraisals.

Specific concerns were raised in TA295 regarding the adjustment that the company used to translate the trial PFS benefit into an OS benefit, stressing the importance of the unresolved uncertainty in PPS. In three further TAs (TA310, TA316, TA333), the ERG and/or the NICE Committee were concerned that the modelled survival results were likely biased by the

application of the hazard ratios from the indirect comparisons or by the external data that were used.

Table 2: Summary of NICE cancer appraisals using the partitioned survival analysis approach

TA no. (STA unless stated)	**Indication**	**Method correctly described**[1]	**Justification provided**[1,2]	**Survival endpoints modelled**	**Differences between treatments for all modelled endpoints**	**External data**	**Adjustment to overall survival**	**PSA**	**Concerns raised by ERG / Committee relating to modelling method**
TA285	Advanced ovarian cancer	Yes	No	2 (PFS, OS)	Yes	No	No	Yes	Uncertain OS extrapolation
TA295	Advanced or metastatic breast cancer	No	Yes	2 (PFS, OS)	Yes	Indirect comparison	Avoid curves crossing	Yes	OS adjustment
TA296	Locally advanced or metastatic non-small-cell lung cancer	Yes	No	2 (PFS, OS)	Yes	Indirect comparison	Treatment switching	Yes	Uncertain OS extrapolation
TA299	Chronic myeloid leukaemia (chronic, accelerated and blast phase)	No	No	2 (TTD, OS)	Yes	Indirect comparison	Surrogacy[3]	Yes	Uncertain OS extrapolation / Treatment effects duration
TA306	Advanced stage non-Hodgkin's B-cell lymphoma	No	Yes	3 (TTD, PFS, OS)	Yes	No	No	Yes	Clinical data robustness
TA307	Metastatic colorectal cancer	No	Yes	3 (TTD, PFS, OS)	Yes	No	No	Yes	Uncertain OS extrapolation / Treatment effect duration
TA309	Advanced or metastatic non-small-cell lung cancer	No	No	2 (PFS, OS)	Yes	No	No	Yes	Treatment effect duration

TA no. (STA unless stated)	**Indication**	**Method correctly described**[1]	**Justification provided**[1,2]	**Survival endpoints modelled**	**Differences between treatments for all modelled endpoints**	**External data**	**Adjustment to overall survival**	**PSA**	**Concerns raised by ERG / Committee relating to modelling method**
TA310	Locally advanced or metastatic non-small-cell lung cancer	Yes	No	2 (PFS, OS)	Yes	Indirect comparison	No	Yes	Uncertain OS extrapolation
TA316	Metastatic prostate cancer	No	No	2 (TTD, OS)	Yes	Indirect comparison	No	Yes	Uncertain OS extrapolation
TA321	Unresectable or metastatic BRAF V600 mutation-positive melanoma	Yes	Yes	2 (PFS, OS)	Yes	Indirect comparison+ Extrapolation	Treatment switching	Yes	OS extrapolation
TA333	Advanced renal cell carcinoma	Yes	No	2 (PFS, OS)	Yes	Indirect comparison	No	Yes	Indirect comparison
TA338	Advanced multiple myeloma	Yes	No	3 (TTD, PFS, OS)	Yes	Indirect comparison	Treatment switching	Yes	Uncertain OS extrapolation
TA344	Chronic lymphocytic leukaemia, all stages	No	No	2 (PFS, OS)	Yes	Indirect comparison+ Extrapolation	No	Yes	Uncertain OS extrapolation / Treatment effect duration
TA347	Locally advanced, metastatic, or locally recurrent non-small-cell lung cancer	Yes	No	2 (PFS, OS)	Yes	Indirect comparison	No	Yes	Uncertain OS extrapolation
TA357	Advanced melanoma	Yes	Yes	2 (PFS, OS)	Yes	Indirect comparison+ Extrapolation	Treatment switching	Yes	Uncertain OS extrapolation

TA no. (STA unless stated)	**Indication**	**Method correctly described**[1]	**Justification provided**[1,2]	**Survival endpoints modelled**	**Differences between treatments for all modelled endpoints**	**External data**	**Adjustment to overall survival**	**PSA**	**Concerns raised by ERG / Committee relating to modelling method**
TA359	Chronic lymphocytic leukaemia, all stages	No	No	3 (TTD, PFS, OS)	Yes	Indirect comparison	Treatment switching	Yes	Treatment effect duration
TA360	Metastatic pancreatic cancer	Yes	No	3 (TTD, PFS, OS)	Yes	Indirect comparison	No	Yes	Uncertain OS extrapolation
TA366	Advanced melanoma	Yes	Yes	2 (PFS, OS)	Yes	Indirect comparison+ Extrapolation	No	Yes	Uncertain OS extrapolation
TA371	Locally advanced or metastatic breast cancer	Yes	No	2 (PFS, OS)	Yes	Indirect comparison	No	Yes	-
TA374 (MTA[4])	Locally advanced or metastatic non-small-cell lung cancer	NR	Yes	2 (PFS, OS)	Yes	No	No	Yes	-
TA378	Advanced gastric cancer	Yes	No	2 (PFS, OS)	Yes	Indirect comparison	No	Yes	-
TA377	Metastatic prostate cancer	No	No	2 (TTD, OS)	Yes	Indirect comparison	Treatment switching	Yes	Uncertain OS extrapolation

TA - technology appraisal; STA – single technology appraisal; MTA – multiple technology appraisal; ERG – Evidence Review Group; PSA – probabilistic sensitivity analysis; OS – overall survival; PFS - progression-free survival; TTD - time to treatment discontinuation; NR – not reported

[1] As per company submission.

[2] Denotes any justification provided beyond precedent.

[3] In TA299, OS for the appraised treatment in one subgroup was estimated using a surrogate relationship linking treatment response to OS. Although the rest of this model follows a PartSA approach, the modelling for this specific treatment in this subgroup is an exception, since survival depends on another endpoint.

[4]The review focused on the AG's model.

4. CRITIQUE OF PARTITIONED SURVIVAL AS AN AID TO DECISION MAKING

Our review of NICE appraisals shows that PartSA is now the most commonly used decision modelling approach in advanced or metastatic cancer. PartSA originated as a method for generating within-trial estimates of quality-adjusted survival, but is now used as a decision modelling approach to provide information about cost-effectiveness to inform policy decisions. This entails using the method to synthesise multiple sources of information, extrapolate parameter estimates beyond the observed data collection period, and quantify the uncertainty around cost-effectiveness estimates. There is therefore a need for consideration of the strengths and limitations of the PartSA approach when used in this context.

4.1. CRITIQUE OF PARTITIONED SURVIVAL APPROACH

4.1.1. Strengths

The strengths of the PartSA approach derive from the direct correspondence between frequently reported time-to-event endpoints such as PFS and OS and the survival functions used within PartSA to derive state membership estimates. This direct correspondence makes the models intuitively appealing, easy to communicate and easy to construct. PartSA directly models each survival curve as a function of time since model entry. This makes it straightforward to reflect any time-dependencies in the event rates (or treatment effects on event rates) corresponding to each survival curve. As the PartSA approach directly models OS, it generally provides accurate predictions of OS for the within-trial period.

PartSA can be implemented using IPD for these commonly reported endpoints, but also by using summary data on these same endpoints. This is an important consideration as accessing IPD for data sources other than the pivotal trial may be difficult if these data are held by competitor companies or parties not directly linked to the appraisal process.

The PartSA approach can be implemented using summary data alone, for example by using methods that allow estimation of IPD from published Kaplan Meier curves.[14, 15] It can also be implemented using a combination of IPD from the pivotal trial alongside external data in summary form. Indeed, in the majority of TAs that used the PartSA approach, IPD from the pivotal trial were used to estimate the required survival functions for comparators included in the pivotal trial. External data in summary form were used to incorporate data on additional comparators using parameter estimates from indirect comparisons, or to inform long-term clinical event rates. Indirect comparisons of cancer treatments commonly provide estimates of hazard ratios for the PFS and OS endpoints. These can be incorporated in to a PartSA model by applying the hazard ratios to the hazard (or cumulative hazard) corresponding to the relevant reference treatment survival curve. When external summary data inform long-term survival functions, the reported statistics may take the form of rates, probabilities, or Kaplan Meier curves, all of which can be incorporated into the PartSA approach.

The direct modelling of OS is also advantageous as the development of some analytic methods has focused on this endpoint. In particular, methods adjusting for treatment switching have focused on generating estimates of OS curves or treatment effects on OS (e.g. hazard ratios). Adjustment for treatment switching was used in six (27%) of the 22 TAs using a PartSA approach. Recommended methods for adjusting for treatment switching include the Rank Preserving Structural Failure Time (RPSFT) method (and related iterative parameter estimation, IPE, method), the Inverse Probability of Censoring Weights (IPCW) method and the two-stage method (see TSD 16 for further details [22]). All methods for adjusting for treatment switching recommended in TSD 16 provide estimates of OS that can be directly incorporated into a PartSA model.

4.1.2. Limitations

The limitations of the PartSA approach stem from its fundamental structural assumption, that the survival functions modelled are independent. Although the conceptual model underpinning PartSA models includes those transitions shown in Figure 3, the implemented structure does not use an explicit disease model and transition probabilities are not

estimated for each possible transition between health states. It is therefore incorrect for applications of the PartSA approach to be described as a form of state transition model as this implies a structural link between health states and that transition probabilities are estimated for each possible transition.

The assumption that the modelled survival endpoints are structurally independent is potentially problematic as there are a number of dependencies between the survival endpoints:

- They include some of the same events (e.g. PFS and OS curves include the same pre-progression deaths);
- Events are structurally dependent (e.g. death cannot be followed by progression and time spent progression-free contributes to time spent alive); and
- Intermediate events are often prognostic for later events (e.g. progression is generally considered prognostic for mortality).

For the within-trial period, these dependencies are reflected in the data and should therefore be closely reflected in the PartSA results. However, for analyses that model beyond the trial period, dependencies between endpoints are ignored with potentially important implications for extrapolation. All TAs we reviewed that used the PartSA approach included some degree of extrapolation and the validity of the extrapolations and the uncertainty they introduced were frequently raised by the ERG and/or Appraisal Committee as concerns.

In a PartSA model, OS extrapolation often reflects the within-trial time trend in the rate of deaths. External empirical evidence or expert judgements can support selection between alternative extrapolation models and evaluation of the plausibility of the extrapolations, or can be more formally incorporated within the survival model.[23] Information on other endpoints, such as the evolving proportion of patients who have progressed, is not explicitly reflected in the extrapolation,[24] though it will have implicitly impacted upon the within-trial mortality trends. This may reduce the validity of the extrapolations produced. For example, consider a single trial control arm, and a simple context in which the hazard (i.e. rate) corresponding to each transition probability depicted in Figure 3 is constant over time. In addition, assume that the risk of death is higher for patients who have progressed than for patients who remain progression-free as progression is prognostic of mortality

risk. One likely scenario is that the cohort of patients who are alive becomes dominated by progressed patients over time. In this context, the hazard of death would initially increase as the proportion of the cohort residing in the progressed state increases, but then level off once all patients had progressed (the hazard levels off at the hazard of death for progressed patients, h_{23}). This is shown in Figure 4(a). In the PartSA approach, extrapolation of the hazard is often based only on the time trend in the hazard observed for the within-trial period (up until time t^*) which is assumed to generalise throughout the extrapolation period. In this example, even with careful parametric model fitting, the PartSA approach may predict that the within-trial trend for an increasing hazard of death with time continues throughout the extrapolation period, as no information on the progression status of patients is taken in to account. This is shown in Figure 4(b). This overestimation of the hazard of death in the extrapolation period will result in an underestimation of survival, and ultimately, an underestimation of mean life years and QALYs. This is shown in Figure 4(c). This very simple case illustrates that by ignoring information on intermediate prognostic endpoints, the PartSA approach can produce inappropriate extrapolations. The risk of such inaccurate projections may be mitigated through the use of external information on long-term hazard trends.

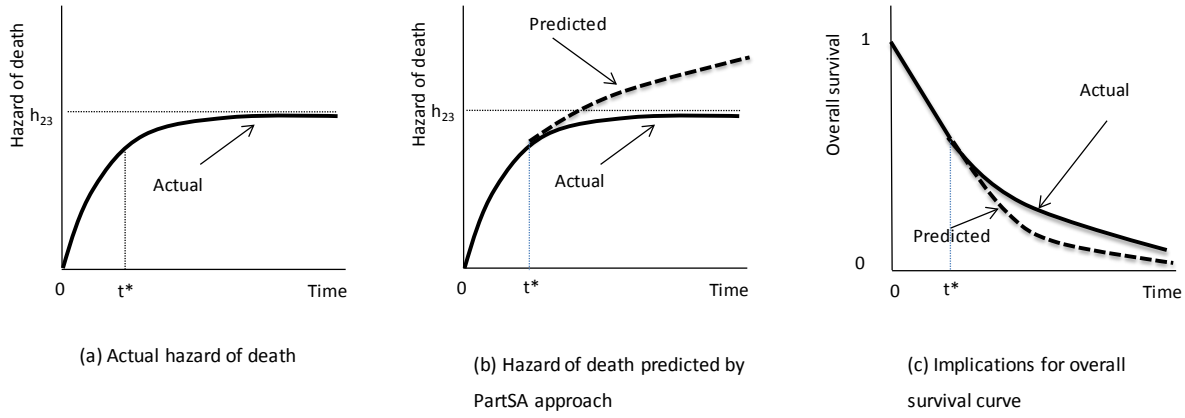
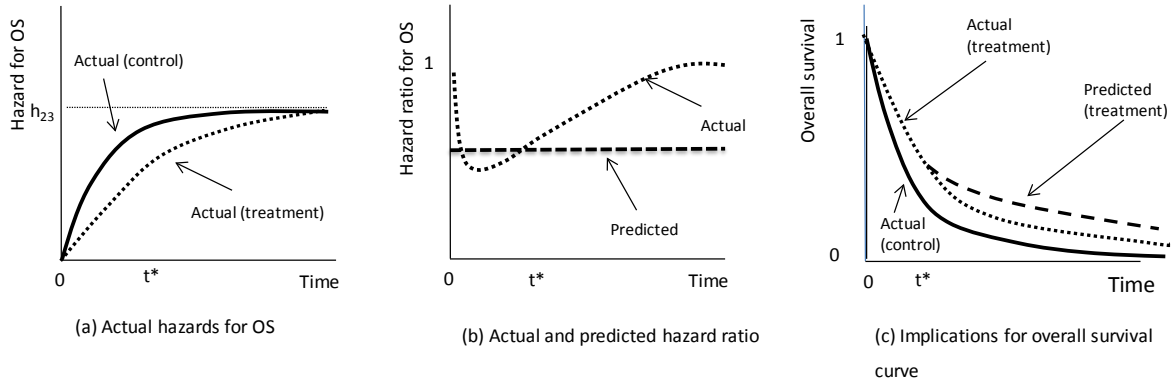


Figure 4: Illustration of potential implications of ignoring information on non-fatal events when extrapolating overall survival

All TAs using PartSA reflected differences across treatments for all modelled survival endpoints. These treatment effects were generally assumed to be constant on the hazard ratio scale or on the acceleration factor scale, for the entire modelled time horizon.

Concerns regarding the extrapolation of constant treatment effects and the implications of this for estimates of PPS were raised by the committee and ERG in a number of appraisals.

These concerns are relevant as it may not be appropriate to extrapolate the observed treatment effect for OS over a lifetime horizon. We consider a case where there is a treatment effect on progression, but not beyond progression. We assume that treatment acts by reducing the hazard of progression (h_{12}) and that the treatment effect on progression conforms to the proportional hazards assumption throughout the model. In our example, differences between treatments in mortality rates are therefore driven solely by differences between treatments in the distribution of patients across prognostic states (i.e. progression-free, progressed). In the earlier part of the model, the treatment arm has a higher proportion of progression-free patients, and therefore a lower hazard of death. If over time, patients in both the treatment and control arms come to reside in the progressed state, the hazard of death in both study arms will converge to the hazard of death for progressed patients (h_{23}). This is shown in Figure 5(a). As the hazard of death in both treatment and control arms becomes similar, the hazard ratio for OS will converge to 1.0. This is shown in Figure 5(b). Depending on the trial time horizon a review of these data may lead to the conclusion that the hazard ratio is decreasing over the long term, approximately stable (i.e. proportional hazards) or increasing over the long term. Ignoring information on the treatment effect mechanism (e.g. that treatments differ in terms of the rate of progression, but not in the rate of death conditional upon progression) and focusing instead on observed time-trends in treatment effects on OS for the within-trial period may therefore result in inappropriate extrapolations of treatment effects. For example, application of a constant hazard ratio i.e. the proportional hazards assumption (as shown by the dashed line in Figure 5(b)) might be deemed plausible based on within-trial trends up to time t* but would result in an over estimation of survival for the treatment arm as shown in Figure 5(c).



(a) Actual hazards for OS

(b) Actual and predicted hazard ratio

(c) Implications for overall survival curve

Figure 5: Illustration of implications of ignoring treatment effect mechanism when extrapolating treatment effects

One way to address concerns regarding the validity of extrapolations generated by PartSA (and other) models is via the use of external data (e.g. longer-term trials, or observational data). These data can be used to inform selection of statistical models or can be explicitly reflected within the model.[16, 23] However, it is often difficult to find external data that are directly relevant to the population and treatments under consideration and external data will in general not be available to inform long-term treatment effects.

Ultimately for the purposes of cost-effectiveness analysis, the focus of any model is on deriving an accurate estimation of differences in mean life years, QALYs and costs between treatments. These differences are determined by both baseline risks and treatment effects. The simplistic examples provided previously illustrate why trends in the within-trial period may not continue in the extrapolation period – either for baseline risk, or the treatment effect. When this is the case, the PartSA method may provide a poor approximation to true incremental survival and QALYs. The true disease process will inevitably be more complex than these simple examples and the extent and direction of differences between the true incremental survival and QALYs and the predictions from the PartSA approach will depend on many factors. These include the number and nature of prognostic states, rates of events and how they change over time, how treatment impacts on these processes, and the maturity of the trial data. The extent to which the PartSA approach over- or under-estimates the true incremental life years and QALYs will be difficult to predict for any specific context, although the extent of differences between predicted and actual outcomes is expected to be less significant as the maturity of the data increases.

The lack of structural relationship between endpoints within the PartSA approach has additional important implications for the appropriateness of using this approach to inform health care decisions. The NICE methods guidance recommends that the "*clinical and biological plausibility*" of extrapolations should be assessed and that alternative scenarios should be routinely considered for the extrapolation period.[8] When decision models are underpinned by a structure reflecting biological or clinical processes, it is possible to carefully consider the mechanisms underpinning extrapolations and to subject these to scrutiny and sensitivity analyses. The structural independence of the modelled endpoints makes this difficult in the context of PartSA. It is possible to review mean time spent within each health state, thus allowing for an assessment of whether extensions to survival are accrued in the progression-free and/or progressed health state, for example. However, it is not possible to review individual transitions which can make the plausibility of extrapolations difficult to assess. For example, if a PartSA model predicts that treatment extends PPS as well as PFS (as observed in a number of the appraisals in our review); this indicates that treatment is associated with reductions in the rate of post-progression mortality. However, as the post-progression mortality transition is not modelled directly in the PartSA approach, it is not possible to establish from the model whether this effect was supported by the trial data in any way or was generated entirely during extrapolation. Within-trial PPS curves could be used to support this assessment (see Section 4.2 for further discussion of issues relating to the interpretation of PPS data). It is also not possible to assess the impact on model predictions of varying the rates of individual transitions, or treatment effects on individual transitions using PartSA models. These sensitivity analyses may be insightful, particularly if there is uncertainty around the nature of any direct or indirect effects of treatment on outcomes post-progression.

The lack of structural relationship between endpoints means that the PartSA approach can predict a PFS curve that lies above the OS curve implying that a higher proportion of patients are progression-free than alive. In the deterministic component of the model, this is most likely to occur in the extrapolation period, though it can occur in the within-trial period if the PFS and OS curves are close to each other and/or the survival models represent a poor fit to the Kaplan-Meier survival curves. A PFS curve that lay above the OS curve was seen in TA295 and led the company to use a form of adjustment to avoid this from happening whereby extensions to PFS were translated to extensions to OS. The

methods used to adjust outcomes were subsequently criticised by the ERG and Appraisal Committee.

The lack of structural relationship between endpoints also reduces the usefulness of PSA results generated using PartSA models. This includes presentations of the distribution of incremental costs and QALYs on the cost-effectiveness plane, presentations of the probability of an intervention being cost-effective such as Cost-Effectiveness Acceptability Curves (CEACs) and estimates of the value of further research. For the within-trial period, this can be addressed by bootstrapping the survival data.[11] For each bootstrap resample, Kaplan-Meier curves or parametric models can then be fitted for all modelled survival endpoints (this was done in only one appraisal in our review, TA321). This will produce a set of correlated survival curves that reflect the relationships between endpoints in the observed data. For the extrapolation period, however, the use of bootstrapping will not reflect the structural relationships between endpoints. This will reduce the validity of the PSA results for the extrapolation period and may also result in logically inconsistent predictions (e.g. PFS higher than OS).

4.1.3. Alternative approaches to specification of survival models for PartSA

There has been some debate about how to specify survival models for the purposes of extrapolation within PartSA models. TSD 14 describes a systematic approach for selecting appropriate survival models to inform cost-effectiveness analyses.[16, 25] The publication of this work triggered some debate regarding alternative approaches to survival model development and selection in this context. Bagust and Beale [24] argued that the validity of extrapolations can be improved by developing statistical models that reflect an understanding of the underlying biological and clinical processes, and the way in which treatment interacts with these processes. They recommend that these considerations should inform the selection of statistical models and that a broad range of models should be considered which may include more complex functions for the baseline hazard, or non-proportional hazards specifications for treatment effects. They advocate using a broader set of information to inform extrapolations, for example, the PPS curves for each treatment may also be reviewed to inform the extrapolation of OS, alongside a range of information on the trial design and anticipated effects of the treatment. One specific approach they

suggest is to identify a time point beyond which there is evidence of the hazard stabilizing at a constant level, and then to focus on this latter part of the data where transient effects are expected to have dissipated. The authors suggest that when such a time point has been identified, a parametric model should be fitted to the remaining data after this cut point to inform the extrapolation. They suggest that the default parametric function should be an exponential (constant hazard) model, unless evidence suggests that this is invalid.

Latimer responded with specific concerns relating to the proposals by Bagust and Beale.[26] Latimer noted that exclusion of data from earlier in the trial represents a loss of information and that the focus of curve fitting on the latter part of the Kaplan Meier curve may introduce additional uncertainty due to the low number of individuals contributing data beyond this point. Latimer also noted the potential for additional subjectivity in the proposed analyses due to the need to identify a cut-point beyond which curve fitting should occur, and raised concerns about the focus on the exponential model in favour of other plausible alternatives.

This debate highlights that there is no single optimal method for extrapolation, and that subjective judgements will inevitably be required to reflect broader considerations and additional information relating to the extrapolation period. This suggests that modelling approaches that are more explicit about the underlying processes and more explicit in the way that additional information is reflected may be warranted.

4.2. ALTERNATIVE APPROACHES

In addition to reviewing NICE appraisals using the PartSA approach, we also reviewed cancer appraisals using alternative approaches. Only eight (27%) of the 30 TAs reviewed used a non-PartSA approach, and all but one of these used a state transition model. The exception was an appraisal in a haematological cancer (TA311) which used a hybrid modelling approach. A decision tree was used to allocate individuals to a treatment response category, and then a PartSA approach was used to determine state membership following achievement of each level of response. The remaining state transition models identified differed from the PartSA models in that they were not generally used as a

vehicle to model the available trial data, but instead to introduce additional assumptions, and in particular to introduce a surrogacy relationship between PFS and OS whereby extensions to PFS resulted in similar extensions to OS. In most appraisals, the justification for this was that the OS data were immature.

The models reviewed tended to make strong assumptions, and frequently assumed that the probability of death from the progressed state ($p(c)_{23}$) was independent of treatment allocation, prior patient history and in some cases time in state. In a number of appraisals, the use of such strong and untested assumptions meant that the models did not predict the within-trial OS satisfactorily. In addition, the analyses that informed the estimation of transition probabilities did not in general use established survival analytic methods for competing and sequential events, and instead used more *ad hoc* approaches. The models were criticised by the ERG and/or Appraisal Committee on the basis that they did not fully utilise available trial data, were not able to predict within-trial data well, used strong assumptions without adequate justification or testing, and did not adequately explore uncertainty. In two cases, the ERG responded to these concerns by re-modelling the decision problem using a PartSA approach, and in these appraisals the Appraisal Committees considered that the PartSA model provided an improved basis for decision making. Further details on the appraisals using non-PartSA approaches are presented in Appendix 3.

The examples and application of alternative approaches (predominantly state transition models) identified in our review do not therefore appear to represent a credible alternative to the PartSA approach. However, the lack of credibility of these alternatives arises largely in terms of their subsequent implementation and specifically the inappropriate exclusion of relevant data and incorporation of untested assumptions. The poor implementation of state transition models historically in TAs of cancer treatments does not provide sufficient grounds for rejecting alternative approaches to PartSA from further consideration.

The predominance of the PartSA approach in cost-effectiveness analyses for cancer treatments differs from other clinical contexts where state transition models are the norm. The advantages of having an underlying biological or clinical process in a decision model are well-established and include that this allows the model to reflect the natural history of a condition (the pathways individuals take during their disease), facilitates a careful

consideration of which health states/events are prognostic, and allows the impact of interventions on that process to be carefully considered.[27] Use of an explicit disease model allows both the natural history of the disease and treatments effects on this to be reflected when extrapolating beyond the trial data, and the assumptions underpinning these extrapolations to be made explicit and therefore subject to scrutiny and sensitivity analyses.

These theoretical and practical advantages extend to cancer, including those contexts in which PartSA models are currently applied. Using a state transition model allows fuller and more explicit use of information on prognostic intermediate endpoints, such as progression, to inform mortality extrapolations. The PartSA approach will only produce reliable predictions to the extent that the OS model used represents changing hazards observed within the trial and can predict how changing health state membership will drive mortality hazards beyond the trial period for each treatment. The state transition approach does this more explicitly by modelling changing health state membership, but will only provide reliable OS extrapolations to the extent that the within trial-trends in the hazard of each event for each treatment continue beyond the trial period. If data are immature the resulting predictions will always be uncertain, regardless of the modelling approach.

State transition modeling allows event rates and treatment effects on these event rates to be specified for individual components of the disease process. As a result, the use of state transition models can improve transparency around the mechanisms and processes underpinning results generated using extrapolation, and facilitate meaningful sensitivity analyses. For new technologies, the pivotal trial(s) will generally provide the sole source of evidence relating to the treatment effect. Any assessment of how treatment effects will evolve beyond the trial period will therefore be subject to considerable uncertainty and is likely to rely upon expert judgments. It may be easier to elicit such judgements for specific transitions. For example, clinical experts may find it more straightforward to consider the effect of treatment on PFS and PPS separately, by reflecting on considerations such as the use of treatments at the pre- and post-progression stages, than on long-term OS which is influenced by multiple processes and lines of treatment.

State transition modelling also avoids logical inconsistencies from occurring which may force adjustments to be made to models using the PartSA approach, as required when PFS and OS curves cross.

There is an established interest in the oncology clinical literature in better understanding the relationships between clinical events to validate surrogate relationships and to more appropriately inform clinical trial design and analysis.[28, 29] Joint modelling of endpoints has been recognised in this context as an important technique to make use of all available information and for generating long-term predictions.[30, 31] The lack of joint modelling of clinical endpoints and predominance of the PartSA approach within existing TAs is therefore perhaps surprising given the established clinical interest in the insights provided by joint modelling.

4.2.1. Implementation of state transition models

The challenge of implementing state transition models in this context is that a series of individual transition probabilities must be appropriately estimated. Where IPD are available for some or all transitions, a set of methods called competing risks analysis and multi-state modelling can be used for this purpose. Competing risks analysis is used when there is a series of "competing" mutually exclusive events (such as progression and pre-progression death). Multi-state modelling is used when events can also occur sequentially (such as progression followed by death). [32 33] A recent tutorial paper by Williams *et al*. [34] provides a step-by-step guide to using multi-state modelling to estimate parameters for state transition models. The methods are illustrated in the context of a previous NICE appraisal in chronic lymphocytic leukemia (TA174), using the model structure outlined in Figure 3.

In both competing risks and multi-state models, individual survival analyses are built for each possible transition, in Figure 3 three separate survival analyses would therefore be necessary. The methods are the same as a standard survival analysis with one adaptation: any event which is not the event of interest is treated as a censoring event i.e. "competing" events are treated in the same way as loss to follow-up.[32] For example, in the analysis informing the transition from progression-free to progressed in Figure 3, pre-progression

deaths are considered to be censoring events. In the analysis of the transition from progression-free to dead, disease progressions are considered to be censoring events. In the analysis of the transition from progressed to dead no additional censoring is introduced, as death is the only possible event. Once this additional censoring has been included, standard survival analysis techniques can be applied using standard software and the full set of parametric models. There are a number of applications of this type of analysis in the cost-effectiveness literature [35] [36] [37] [38] though none that we are aware of in an advanced or metastatic cancer setting. In the wider clinical literature multi-state modelling (which when used in the context of a three state model is often referred to as "illness-death" modelling) has been applied in both early- and late-stage oncology settings to explore the implications of intermediate events for long-term prognosis and to improve understanding of the effects of treatment.[29, 39-42] Further methods are available which allow for different levels of data availability.[43] [44] Methods that address the fact that the exact date of clinical progression is often not known (i.e. data are interval censored) may be relevant.

Each transition probability can vary in different ways over time (e.g. via the use of different parametric models) and different treatment and covariate effects can be applied to each transition. Estimation of the transition probabilities therefore requires careful consideration of the survival model for each transition. Guidance regarding survival model selection is available for individual and independently modelled survival endpoints, as required to parameterise PartSA models.[16, 24] However, little guidance is available to inform survival model selection in the multi-state setting, particularly in the context of cost-effectiveness analysis where extrapolation is the primary purpose of the statistical modeling. Assessing model fit is more challenging in this context as the target quantities of interest such as proportions of individuals experiencing each event, are no longer determined by a single survival model, but are instead determined by a combination of survival models. The Williams *et al.* tutorial provides some guidance and suggests comparisons of predicted and observed proportions of patients in each health state over time, use of log cumulative hazard plots, use of cumulative hazard plots and use of general model fit statistics. Titman and Sharples discuss methods for diagnosing model fit in the context of Markov models,[45] though they acknowledge that extensions would be required to translate these to semi-Markov models. The limited guidance regarding model selection, particularly in the context of cost-effectiveness analysis, and wide set of choices available [40] increases the likelihood that the methods could be used inappropriately. In addition, the

validity of the choices made will be difficult to validate for ERGs and AGs who do not typically have access to the underlying IPD.

While there are several theoretical and practical advantages of state transition models, there also remain some important issues and challenges relating to their implementation when data are available on a series of interlinked endpoints, including: (i) the availability of data in a format that allows all required parameters to be estimated; (ii) use of methods to estimate transition probabilities that provide a plausible basis for extrapolation; and (iii) implementation of time-dependent transition probabilities within state transition models.

(i) Data availability. Use of state transition modelling, parameterised by using multi-state survival analysis to estimate transition probabilities, is feasible when IPD are available. IPD were available for the pivotal trial in almost every reviewed NICE appraisal that used PartSA. However, the external data that informs indirect comparisons and long-term outcomes is typically only available in summary form.

As shown by the review of cancer appraisals, parameters estimated from indirect comparisons are often used to reflect the impact of interventions on progression and mortality within cost-effectiveness modelling. If survival curves can be obtained for time to progression (pre-progression deaths censored), time to death without prior progression (progressions censored) and PPS, then the IPD corresponding to each transition in Figure 3 can be recovered. Where survival curves are available for PFS, OS and PPS, it is also possible to recover this information, though this requires either knowledge of whether PFS events were progressions or deaths, or an assumption that events occurring simultaneously within the PFS and OS curves were pre-progression deaths (as assumed by Williams *et al*. [34]). Under these scenarios all transition probabilities can be estimated, although without access to the underlying IPD it is not possible to quantify the impact of patient-level covariates on PPS which may be important for extrapolation (see issue (ii) below).[34]

Once time-to-event data corresponding to each transition have been derived, evidence synthesis can be conducted. For the model shown in Figure 3, synthesis of relative treatment effects would be required for each individual transition. For the PPS transition ($p(c)_{23}$) standard methods could be applied. However, for competing events (e.g. $p(c)_{12}$ and $p(c)_{13}$) more advanced methods (i.e. competing risks network meta-analysis) are required

to capture the negative correlation between outcomes.[46] To date, applications of competing risks network meta-analysis have assumed constant baseline hazards and proportional hazards between treatments and between competing events. Extensions to these methods would be required to increase their flexibility.

In many instances, external data are only available for the PFS and OS endpoints. These data are not sufficient to allow estimation of individual transition probabilities. A range of methods could be applied in this context. For example, an analysis of the pivotal trial could provide transition probability estimates for those treatments within the pivotal trial. This would allow PFS and OS predictions to be generated for these treatments, to which external estimates of treatment effects could be applied. Alternatively, calibration or multiple parameter evidence synthesis could be used to effectively calibrate the analysis of the pivotal trial to replicate the PFS and OS outcomes for external trials. However, we are not aware of any research to date on this issue.

These difficulties could be reduced if data were available on the individual transitions for all relevant clinical data sources. In the longer term, trialists and others reporting time-to-event data should be encouraged to report individual endpoint outcomes and to share IPD.

We have focused on indirect comparisons here as this is the context in which external data was most commonly used in the reviewed appraisals; however the considerations outlined are similar to those faced when using external data to inform long-term predictions.

(ii) Generalisability of transition probability estimates. A second set of considerations relates specifically to estimating transition probabilities from health states other than the original entry health state, for example the probability of death from the progressed state, $p(c)_{23,}$ in Figure 3. Estimation of these transition probabilities introduces three issues.

The first is a selection issue and can be illustrated by considering estimation of $p(c)_{23}$. Only those patients who have experienced a progression event within the trial follow-up will inform estimation of $p(c)_{23}$. This could introduce bias in the extrapolation period if patients who progress within the trial are not representative of those who progress later. To illustrate the implications of this we consider an example in which patients who have a better prognosis spend a longer period progression-free and a longer period with

progression, and treatment delays progression. In this example, a larger proportion of people will have progressed in the control arm than in the treatment arm within the trial follow-up. Those who have progressed in the control arm will have a better prognosis on average than those who have progressed in the treatment arm. A naïve comparison of the PPS curves in this instance would suggest that patients in the treatment arm fair worse after progression than those in the control arm, despite the fact that there is no effect of treatment on PPS. In addition, a naïve analysis of PPS will underestimate PPS in both trial arms as the PPS data relates to "early progressors". These selection effects have been recognised in the literature on cancer endpoints.[47] Selection effects should be carefully considered whenever PPS data are used and naïve comparisons and extrapolations of PPS should be viewed with caution. Guidance regarding addressing selection effects is available in TSD 17,[48] though this focuses on the estimation of treatment effect parameters only. Further research is required to consider the appropriate application of such methods to estimating transitions from intermediate health states where the primary objective is to inform extrapolation, and there may be a desire to preserve some differences in patient characteristics across study arms at the point of progression.

The second issue relates to the underlying assumption typically made in survival analysis that censoring is "independent" i.e. conditional upon any covariates included in a survival analysis, at each time point the censored individuals are representative of those individuals remaining at risk of an event.[49] In a clinical trial setting patients who progress later will tend to be censored earlier for PPS due to administrative censoring associated with the end of trial follow. If time taken to progress is prognostic then this will mean that censoring is dependent which will bias the estimation of transition probabilities. Again, consider the example provided above where patients who take longer to progress also have a lower risk of death post-progression. Towards the tail of the PPS curve early progressors will be over-represented amongst the at-risk population which will result in an overestimation of the rate of death. The potential for dependent censoring to bias results has been discussed in the statistical literature on multi-state modelling/event history analysis.[50] The potential for this bias to occur should be considered and where appropriate, adjusted for.

A third set of challenges relate to the ability of the survival analyses underpinning state transition models to provide a reasonable fit to the observed within-trial data. Achieving a satisfactory fit to the observed data is more challenging within state transition models than

within PartSA models. This is because the target quantities of interest (e.g. OS) are no longer determined by a single survival model, but are instead determined by a combination of survival models which may also be related to each other (e.g. PPS may depend on time taken to progress), and transition probabilities may be subject to selection biases and dependent censoring.

Further research is required to develop methods for estimating transition probabilities that appropriately account for selection issues and dependent censoring, and credibly represent the within-trial data.

In addition, depending on the nature of the decision problem and the characteristics of the available clinical data, there may be other estimation challenges. For example, one of the methods for adjusting for treatment switching recommended in TSD 16 [21] – the IPCW method - currently produces outputs in formats which cannot be readily used to derive transition probabilities for a state transition model – though RPSFT and two-stage methods could be used to inform transition probability estimates.

(iii) Implementing time-dependent transition probabilities. Finally, where the transition probability estimates depend on time spent in an intermediate health state or time in previous states, appropriate methods are required to implement the state transition model. These include use of tunnel states,[27] semi-Markov approaches,[51] and patient level simulation.[19]

It may also be possible to retain a simple structure and a cohort simulation approach whilst incorporating time-dependency in transition rates from intermediate states, via the use of a "payoff" approach. This has been explored in an applied setting[52] and involves modelling transitions from the entry state (e.g. $p(c)_{12}$, $p(c)_{13}$) in a standard way, but then, rather than explicitly modelling the transitions experienced by patients from the progressed state via $p(c)_{23}$, instead assigning these individuals a "payoff" representing their total post-progression costs and QALYs. This payoff can be calculated using the area under the PPS curve to estimate time spent in the progressed state and then applying costs, utilities and discounting to this time period. Care must be taken to discount costs and QALYs at the time point at which they actually accrue. Although the "payoff" approach uses the AUC method to estimate state membership for the progressed state, it does not represent a

PartSA method because it involves survival curves that are mutually exclusive. Instead it is a method that makes exactly the same assumptions as a standard state transition model, but uses AUC calculations to simplify implementation. This allows $p(c)_{23}$ to be time-dependent whilst avoiding the need to track the time patients have spent in an intermediate health state via tunnel states or simulation methods. The drawback of this approach is that it does not allow model predictions, such as OS curves, to be generated. This method could potentially be used to improve computation (e.g. for probabilistic and other sensitivity analyses) alongside a patient-level simulation or complex state transition model which would allow projected survival to be produced for the base case and other key scenarios.

Even where probabilities of transitioning from an intermediate state appear to be constant over time, if it is considered important to explore the sensitivity of the model to alternative parametric survival functions, then increased flexibility and complexity will need to be built in to models from the outset.

Care needs to be taken to select between these alternative methods for implementing state transition models, further guidance on this issue is available in TSD 15.[19] This will result in decision models that are more complex than the PartSA models, and that take longer to develop, validate and review.

4.3. EMPIRICAL COMPARISONS OF PARTSA AND STATE TRANSITION MODELLING

Four studies have attempted to compare PartSA and state transition models empirically. These are the published work by Goeree *et al.*[53] and Williams *et al.*[38] and conference presentations by Coyle and Coyle [54] and Briggs et al.[55]

The work by Goeree *et al.* compares a PartSA model to a Markov model in an advanced cancer setting. The approach used for the Markov model is unclear and does not appear to include any conditional transition probabilities (e.g. there is no reported estimate of mortality from the progressed state). We do not therefore consider that this study represents a comparison of state transition and PartSA models.

The work by Williams *et al*. compares the PartSA approach to a state transition model in which transition probabilities were estimated using multi-state modelling in the context of a cost-effectiveness analysis of rituximab for chronic lymphocytic leukaemia. The model structure was as shown in Figure 1 for the PartSA approach and Figure 3 for the state transition model. The authors found that in this example, the approaches differed substantively in their state membership predictions in the extrapolation period, and that the state transition model predicted a lower incremental QALY gain than the PartSA approach. The drivers of this difference are not entirely clear, and it is possible that different approaches to the selection of the survival models may have influenced the results. Nonetheless, one important driver is likely to have been the nature of the treatment effects estimated and applied within each model. In the PartSA model constant hazard ratios of 0.6 and 0.8 were applied throughout the model time horizon to PFS and OS respectively. In the state transition model, constant hazard ratios of 0.6, 0.7 and 1.4 were applied to transition probabilities $p(c)_{12}$, $p(c)_{13}$, and $p(c)_{23}$ respectively. As the alive cohort in the state transition model became more dominated by progressed patients, the negative treatment effect on PPS (as indicated by the hazard ratio exceeding 1.0 for $p(c)_{23}$) resulted in convergence of the OS curves. This was not observed within the PartSA model due to the proportional hazards assumption on the OS endpoint. This resulted in incremental life year estimates for the state transition model that were approximately half those predicted for the PartSA model. This example shows that the different modelling approaches may produce markedly different results, though it is not clear which is more reliable, particularly as the analysis of PPS in Williams *et al.* did not discuss or address the potential for selection effects or dependent censoring.

Limited details were available for the other studies. The work by Coyle and Coyle uses a simulated dataset to evaluate the predictions made by the PartSA approach. The authors conclude that the PartSA method is subject to bias based on their simulation though few details were available. Finally, the work by Briggs *et al*. compared a partitioned survival model to a state transition model in an advanced cancer setting. The approaches were reported to give similar results though again, few details were available.

5. SUMMARY

PartSA represents a specific type of decision model with the defining feature that state membership is determined by a set of independently modelled non-mutually exclusive survival curves, that model – for each alive health state - time from model start (i.e. patient entry in to the model) to transiting to *any health state that is further along the sequence.*. This distinguishes PartSA from state transition models, which explicitly link the modelled clinical events. In the context of a within-trial analysis or a case in which data have been fully observed, PartSA and state transition modelling are expected to produce similar results as relationships between endpoints are reflected within the data. When there is a need to extrapolate parameter estimates beyond the observed data collection period, the approaches make different assumptions and use different information, and are therefore expected to produce different results. In particular, extrapolated mortality in a PartSA model is determined by prior trends in mortality rates whereas extrapolated mortality in a state transition model is determined by a structural link between mortality and earlier disease-related events such as cancer progression. The effects of treatments on disease processes are also specified differently. In PartSA treatment effects act directly on the survival curves (e.g. PFS, OS) whereas in state transition models treatment effects can act on different parts of the disease process and the interaction of these effects will determine the overall impact on the estimated survival curves. PartSA models can be directly parameterised using commonly reported survival endpoints (e.g. PFS, OS) whereas state transition models require estimates of individual transition probabilities. A summary of the attributes of PartSA in contrast to state transition models is provided as Table 3.

	Partitioned survival analysis	State transition modelling
Basis for estimating state membership	Set of survival curves describing state membership across non-mutually exclusive groups of health states. Basic manipulations used to derive health state membership for individual health states	Transition probabilities for each possible transition used to describe disease progression. Simulation methods (cohort or patient-level) used to derive state membership
Data inputs	Time to event data for routinely reported clinical endpoints (e.g. PFS, OS), derived from summary or IPD	Time to event data on individual transitions This may not be available, particularly for external data required to inform indirect comparisons* Difficult to validate or verify without access to IPD
Methods for reflecting time-dependency in event risks	Time-dependency in risks underlying survival curves can be reflected directly	Implementation of time-dependent transition probabilities may require use of tunnel states, semi-Markov models, patient level simulation or "payoff" approaches
Basis for extrapolation of overall survival	Time trends in OS risk	Rate of progression between health states and mortality risk conditional upon health states
Basis for extrapolation of treatment effects on overall survival	Time trends in the treatment effect on overall survival	Treatment effects on individual transitions jointly determine treatment effect on overall survival
Risks to validity of extrapolations of OS	Trends in rates of death, and treatment effects on rates of death may not translate to extrapolation period	Naive estimation of outcomes from intermediate health states (e.g. PPS) may not be reliable.* Clearer link between health state membership mix and long-term event trends, but the risk remains that trends in rates of individual transitions, and treatment effects on rates of individual transitions may not translate to extrapolation period
Considerations for use within decision making process	More difficult to assess plausibility of extrapolations More difficult to subject to meaningful sensitivity analyses	Easier to assess plausibility of extrapolations Easier to conduct meaningful sensitivity analyses

Table 3: Comparison of attributes of partitioned survival analysis and state transition models

* Denotes areas in which further methods research is warranted. IPD – Individual patient data; OS – Overall survival; PFS – Progression-free survival; PPS – Post-progression survival.

5.1. USE OF PARTSA IN NICE TECHNOLOGY APPRAISALS

Our review of 30 recent oncology TAs found that PartSA was used in 73% of the appraisals. The method was generally described poorly and inaccurately in these appraisals, and little attention has been paid to its underlying assumptions or relative merits compared to alternative approaches. The most common PartSA structure used two survival endpoints (OS, and either PFS or TTD) to partition individuals into three health states and almost all appraisals applied continuous treatment effects to each modelled survival endpoint for the full model time horizon. The majority of the appraisals included a naïve or adjusted indirect comparison within the modelling to facilitate inclusion of additional comparators. This generally required use of data external to the pivotal trial. The most common concerns raised by ERGs and committees during the appraisals related to the uncertainty around modelled long-term OS projections. These included concerns relating to implausible predictions (e.g. where PFS was observed to exceed OS at specific time points) and model predictions that suggested that the appraised treatment extended PPS.

5.2. ADVANTAGES AND DISADVANTAGES OF PARTSA

The direct correspondence between frequently reported time-to-event endpoints such as PFS and OS and the survival functions that inform state membership estimates in PartSA makes the models intuitively appealing, easy to communicate and construct, allows replication of the within-trial data with relative ease, and means that PartSA models can be constructed using either summary data or IPD for these endpoints. This is particularly important as for data sources other than the pivotal trial access to IPD may not be possible.

The limitations of the PartSA approach stem from its fundamental structural assumption, that the survival endpoints are independent. This has a number of implications. In general, extrapolations for a given endpoint reflect within-trial trends in that endpoint alone. Using simple scenarios, it is possible to show that extrapolating within trial trends in baseline risk and treatment effect in this way, without considering the underlying disease process, may not produce appropriate extrapolations. Some work has attempted to overcome this limitation via more considered development of statistical models for extrapolation;

however the proposed methods increase the role of subjective decisions and only informally reflect the full set of available information. The lack of a link between clinical endpoints also limits the degree to which the biological and clinical plausibility of extrapolations generated by the PartSA approach can be subject to scrutiny and sensitivity analyses. Finally, assuming independence between endpoints reduces the value of PSA as a means of quantifying decision uncertainty.

5.3. ALTERNATIVES TO PARTSA

Almost all of the appraisals reviewed that did not use the PartSA approach used a state transition modeling approach. The state transition models reviewed made strong assumptions without adequate justification and often excluded relevant data. This poor implementation of state transitions models does not provide sufficient grounds for rejecting alternative approaches to PartSA from consideration.

State transition models incorporate an explicit link between clinical endpoints. This means that extrapolations depend upon state membership at the end of trial follow-up, the model structure, and within-trial estimates of each transition probability. This allows the prognostic nature of intermediate health states to be reflected in the extrapolation period, and differential treatment effects to be applied to different components of the disease process. It also allows the processes driving extrapolated results to be reviewed and subject to sensitivity analysis.

The main challenge in the use of state transition models relates to the estimation of the required transition probabilities. A set of survival analytic methods is available for this purpose called competing risks and multi-state modelling. However, there remain a series of challenges in applying these methods to develop cost-effectiveness models in cancer. Summary data play an important role in many cancer appraisals in informing indirect comparisons and long-term extrapolations. In many cases, the available summary data may not be sufficient to facilitate estimation of individual transition probabilities using existing methods. Simple analyses of time-to-event data from non-entry health states (e.g. PPS) may be subject to biases due to selection effects and informative censoring. Achieving a satisfactory fit to the observed within-trial data may be challenging as the target quantities

of interest (e.g. OS) are no longer determined by a single survival model, but are instead determined by the combined effect of all transition probability estimates. In addition, whilst the state transition approach allows a clearer link between the health state membership and trends in rates of death and treatment effects on rates of death, the risk remains that trends observed in the trial period for individual transitions may not translate to the extrapolation period and that extrapolations may therefore be unreliable. Finally, the application of state transition models to the settings in which PartSA is currently used may necessitate the use of more complex approaches to model implementation (e.g. semi-Markov models or individual-level models) to appropriately reflect time-dependencies in event rates.

5.4. AREAS FOR FURTHER RESEARCH

There is a need for research focused on the practical application of state transition modelling to the settings in which PartSA is currently used. The first set of research relates to the estimation of transition probabilities using appropriate statistical modelling that takes in to account the competing and sequential nature of the events modelled, the need for careful model selection, and the need for careful consideration of, and adjustment for, potential biases introduced by selection effects and dependent censoring. The second set of research relates to the challenge of estimating individual transition probabilities when only summary data are available, as summary data are typically reported for the survival endpoints modelled by PartSA rather than the individual transition probabilities required by state transition models. Further research to understand the conditions under which the PartSA and state transition modelling approaches perform well is also warranted.

6. RECOMMENDATIONS

6.1. SELECTING, DOCUMENTING AND JUSTIFYING THE MODELLING APPROACH

In the NICE TAs using the PartSA approach, the rationale for selecting the modelling approach, and an explicit summary and justification of the main structural assumptions were not routinely reported.

Recommendation 1: The model conceptualisation process should be routinely reported and the rationale for the chosen modelling approach explicitly justified on the basis of theoretical and practical considerations. A summary of the model conceptualisation process should be routinely reported in submissions (see TSD 13 for further discussion of the model conceptualisation process).[56] This should document the theoretical and practical considerations which support the chosen modelling approach and structure. The justification for choice of modelling approach and structure should be more closely aligned to the key features of the disease/technology. These include: the health-related events that occur over time, whether and how the risks of events change with time and with the occurrence of other health-related events, and how treatment modifies these events and over what period.[27, 57] Given the need for extrapolation, and the potential significance of extrapolations in determining cost-effectiveness, clearer consideration of how the modelling approach may support appropriate extrapolations is required. The choice of modelling approach may be constrained by the available evidence. In these circumstances, this restriction should be formally stated and any limitations should be clearly acknowledged. The PartSA approach is intuitive, easy to implement and generally predicts trial endpoints well for the within-trial period. However, this does not provide a sufficient basis for justifying the PartSA as an appropriate modelling approach for informing decisions relating to cost-effectiveness. It is important to recognise that PartSA may not provide the ideal modelling approach to inform extrapolation due to the lack of structural relationship between modelled endpoints.

Recommendation 2: Consistent and appropriate terminology should be applied in future appraisals when describing the PartSA approach (e.g. use of the term

"Partitioned survival analysis"). PartSA is often described inaccurately. The terminology and associated description of the PartSA approach should be more clearly distinguished from state transition modelling (e.g. Markov and semi-Markov approaches) and references to these should be avoided to prevent potential confusion e.g. the PartSA approach should not be described as similar to a Markov approach or as 'Markov like'.

Recommendation 3: A summary of the main structural assumptions should be routinely reported and justified as required by the NICE guide to the methods of technology appraisal.[8] For PartSA models this should include a clear statement of and justification for, the following structural assumptions (if applicable): (i) all endpoints - including OS - are modelled and extrapolated independently; and (ii) trends in the hazard of each endpoint and treatment effects on these hazards observed within the trial period are assumed to generalise to the extrapolation period. For state transition models this should cover allowed transitions, the transitions on which differences between treatments are assumed to occur due to direct or indirect effects of treatment, any assumptions about the (in)dependence of post-baseline transitions on prior patient history, and any surrogacy relationships implied by the nature of the treatment effects.

6.2. REPRESENTING UNCERTAINTIES ASSOCIATED WITH EXTRAPOLATION

The structural independence between endpoints in the PartSA approach means that extrapolations for a given endpoint will in general reflect within-trial trends *in that endpoint alone*. Extrapolating within-trial trends without considering the underlying disease process may not produce appropriate extrapolations.

Recommendation 4: All stakeholders should recognise the specific limitations of PartSA for the purposes of extrapolation. All methods of extrapolation are subject to uncertainty, whether using a PartSA approach or a state transition approach. However, the failure to take account of information on intermediate endpoints in the PartSA approach may increase the uncertainty associated with the extrapolations generated using this method. The lack of explicit disease processes underpinning the extrapolations from PartSA models also limits the possibilities for assessing their credibility. In addition, whilst it is straightforward to explore some scenarios in a PartSA framework (for instance, using

different parametric survival functions, or applying equal hazard rates between treatment groups after a set time-point), others are more difficult to explore – such as the relationship between the treatment effect and specific health states.

Recommendation 5: Modelling choices that influence outcomes in the extrapolation period should reflect all relevant evidence. All relevant evidence should be considered and reported when informing the specification of models for the extrapolation period, and this should not be based solely on statistical considerations. Relevant evidence may include data external to the pivotal trial and further information from within the pivotal trial (see Recommendation 6). External data can be used to assess the plausibility of the extrapolated portions of parametric survival models. These data may provide information on outcomes for untreated patients or patients treated with established interventions. It is unlikely that empirical data will be available with which to assess the plausibility of long-term outcomes for newer therapies, and assessment of plausible trajectories for treatment effects beyond the trial period is likely to rely heavily upon expert reasoning and judgements.

Recommendation 6: Within-trial survival curves corresponding to individual clinical events should be supplied alongside PartSA models. Submissions typically report survival curves for commonly reported endpoints such as PFS and OS for the pivotal trial. These should be accompanied by hazard plots to demonstrate whether changing trends over time can be observed, both with respect to survival and with respect to the treatment effect. This may provide useful information on whether hazards and treatment effects appear to be changing as health state membership changes. In addition, survival curves for individual clinical events should be generated and reported, to provide additional information on the underlying processes. For example, for a three-state progression-free, progressed, dead model, information should be provided on time to progression (with deaths censored); time to death without prior progression (with progressions censored) and PPS. This information should be provided in the form of Kaplan-Meier curves (including numbers at risk) for each treatment. Considerations when interpreting these data are discussed in Recommendation 7.

Recommendation 7: When extrapolation of the trial evidence is required to appropriately inform cost-effectiveness, PartSA models should easily facilitate the investigation of alternative assumptions in accordance with current NICE methods

guidance.[8] This should include reflection of uncertainty in baseline risk and treatment effects over the extrapolation period. Scenarios for the extrapolation period should reflect the range of outcomes deemed plausible based on all relevant data, including data external to the pivotal trial, and data on individual clinical events within the trial (see Recommendations 5 and 6). In addition, for treatment effects, scenarios suggested by the NICE methods guidance for the treatment benefit in the extrapolated phase should be explored, namely: (i) nil; (ii) the same as during the treatment phase and continues at the same level; or (iii) diminishes in the long term. The reliance of conclusions relating to cost-effectiveness on specific extrapolation assumptions should be clearly acknowledged by all stakeholders.

6.3. USE OF ALTERNATIVE MODELLING APPROACHES

The lack of structural link between endpoints in PartSA models may increase the potential for inappropriate extrapolation, and may make it difficult to understand the mechanisms underpinning extrapolations and therefore to assess their clinical and biological plausibility. In addition, within a PartSA framework it is difficult to explore informative alternative scenarios around the relationship between event rates and specific health states, and between the treatment effect and specific health states. In circumstances where intermediate health states are known to be prognostic or indicative of a change in treatment effect, and extrapolation is required, state transition models may overcome these issues and therefore confer theoretical and practical advantages over PartSA approaches.

State transition modelling has been implemented poorly in the past – making strong and untested assumptions that may not have credibly represented the observed data, nor generated credible extrapolations. There is an emerging literature that aims to apply these methods in a more robust way.

Recommendation 8: Transition probabilities within state transition models should be estimated using appropriate statistical methods and reflect all relevant evidence. State transition models should use established multi-state modelling techniques to estimate transition probabilities where possible, clearly justify and evaluate any assumptions made, and carefully consider the generalisability of parameter estimates to the extrapolation

period. Biases caused by selection effects and informative censoring should be considered and where possible adjusted for when estimating transition probabilities. Predictions generated by state transition models should be verified against the observed endpoint data for the within-trial period. In addition, there should be some consideration of how external data may support the specification of models of the individual transitions. Given the lack of examples and established methodology for incorporating the results of indirect comparisons based upon summary data in state transition models, any analyses attempting to do this should be examined carefully and any assumptions required made clear. ERGs and AGs may require access to additional summary data and potentially IPD in order to validate state transition models. In addition to any statistical modelling, the Kaplan Meier curves corresponding to each transition should also be presented, as outlined in Recommendation 6.

Recommendation 9: Further research and guidance is required to support appropriate specification of state transition models using multi-state survival analysis. Some recent work has shown the feasibility of using multi-state survival analysis to inform state transition models in this setting, however there remain a number of fundamental questions regarding how to specify the required survival models. The first relates to how post-baseline transition probabilities, and treatment effects, can be robustly estimated in the presence of censoring. This is a critical issue as such post-baseline transitions (e.g. PPS) are often key drivers of OS extrapolations. The second relates to how model fit should be assessed given that typical model outputs to which a good fit is expected (e.g. OS) are a function of multiple fitted survival functions. Thirdly, further research is required to support the development of credible model structures, and in particular to support the selection of intermediate endpoints that represent credible markers for changes in prognosis and/or treatment effects.

Recommendation 10: Further research is required to support incorporation of data external to the pivotal trial, and in particular data used to inform indirect comparisons, in to state transition models. This research should focus on the typical situation faced in TAs whereby IPD is available for the pivotal trial(s) but not for comparator trials, for which data is most likely to be available in the form of aggregate survival curves for key endpoints (e.g. PFS, OS).

Recommendation 11: State transition modelling should be used alongside the PartSA approach to assist in verifying the plausibility of PartSA's extrapolations and to address uncertainties in the extrapolation period, even if this is only plausible for the pivotal trial. A strong recommendation in favour of completely replacing PartSA with state transition models cannot be made, given the need for further research (see Recommendations 9 and 10). This recommendation should be reviewed as further research on this topic emerges. Whilst conferring some advantages over PartSA with respect to extrapolation because the changing health state occupancy mix is explicitly modelled, stakeholders should note that the validity of the extrapolation remains dependent upon within trial trends in individual transition rates being representative of post-trial trends.

Recommendation 13: Presentation of results from all PartSA and state transition models should include tabulations showing the states in which life year and QALY differences between interventions accrue and a justification of why these differences should be considered plausible. The NICE STA evidence submission template requires these tabulations. These tabulations should be accompanied by both an explanation of the mechanism by which the model generated the observed differences, and a justification for why they are plausible based upon available evidence.

Recommendation 14: Further research is required to identify the extent of possible biases associated with PartSA and state transition models, and how this varies according to the context in which the approaches are used. This could comprise a simulation exploring a variety of different scenarios about how the risks of events vary over time, how they relate to intermediate states and patient history; how treatment impacts upon the event rates; and the degree of censoring in the empirical time-to-event data. This research should explore the extent to which implementation of PartSA using more flexible parametric survival models could impact upon the reliability of the resulting extrapolations.

REFERENCES

1. Sonnenberg F, Beck R. Markov Models in Medical Decision Making: A Practical Guide. Medical Decision Making. 1993;13(4):322-38.
2. Barton P, Stirling B, Robinson S. Modelling in the economic evaluation of health care: selecting the appropriate approach. Journal of Health Services Research & Policy. 2004;9(2):110–8.
3. Brennan A, Chickb S, Davies R. A taxonomy of model structures for economic evaluation of health technologies. Health Economics. 2006;15:1295–310.
4. Hoyle M, Peters J, Crathorne L, et al. Cost-Effectiveness of Cetuximab, Cetuximab Plus Irinotecan, and Panitumumab for Third and Further Lines of Treatment for KRAS Wild-Type Patients with Metastatic Colorectal Cancer Value in Health. 2013;16:288–96.
5. Hornberger J, Rickert J, Dhawan R, Liwing J, Aschan J, Lothgren M. The cost-effectiveness of bortezomib in relapsed/refractory multiple myeloma: Swedish perspective. European Journal of Haematology. 2010;85:484–91.
6. Delea T, Amdahl J, Nakhaipour H, Manson S, Wang A, Chit A. Cost-effectiveness of pazopanib in advanced soft-tissue sarcoma in Canada. Current Oncology. 2014;21:748-59.
7. Rochau U, Jahn B, Qerimi V, et al. Decision-analytic modeling studies: An overview for clinicians usingmultiple myeloma as an example. Critical Reviews in Oncology/Hematology. 2015(94):164–78.
8. National Institute of Health and Care Excellence, Guide to the methods of technology appraisal. 2013.
9. Siebert U, Alagoz O, Bayoumi A, et al. State-transition modeling: a report of the ISPOR-SMDM modeling good research practices task force-3. Value in Health. 2012;15(6):812-20.
10. Pazopanib for the first-line treatment of patients with advanced renal cell carcinoma (RCC). Submission from GlaxoSmithKline UK. 2010;National Institute For Health And Clinical Excellence:1-315.
11. Glasziou P, Simes R, Gelber R. Quality adjusted survival analysis. Statistics in Medicine. 1990;9:1259-76.
12. Goldhirsch A, Gelber R, Simes J, Glasziou P, Coates A. Costs and benefits of adjuvant therapy in breast cancer: a quality-adjusted survival analysis. Journal of Clinical Oncology. 1989;7(1):36-44.
13. Tate W, Skrepnek G. Quality-adjusted time without symptoms or toxicity (Q-TWiST): patient-reported outcome or mathematical model? A systematic review in cancer. Psycho-Oncology. 2015;24:253–61.
14. Guyot P, Ades A, Ouwens M, Welton N. Enhanced secondary analysis of survival data: reconstructing the data from published Kaplan-Meier survival curves. BMC Medical Research Methodology. 2012;12(9).

15. Hoyle M, Henley W. Improved curve fits to summary survival data: application to economic evaluation of health technologies. BMC Medical Research Methodology. 2011;11(139).
16. Latimer N. Survival Analysis for Economic Evaluations Alongside Clinical Trials - Extrapolation with Patient-Level Data: Inconsistencies, Limitations, and a Practical Guide. Medical Decision Making. 2013;33:743-54.
17. Jansen J. Network meta-analysis of survival data with fractional polynomials. BMC Medical Research Methodology. 2011;11(61).
18. Lu G, Ades A, Sutton A, Cooper N, Briggs A, Caldwell D. Meta-analysis of mixed treatment comparisons at multiple follow-up times. Statistics in Medicine. 2007;26(10):3681-99.
19. Davis S, Stevenson M, Tappenden P, Wailoo A. NICE DSU Technical Support Document 15: Cost-Effectiveness Modelling Using Patient-Level Simulation. NICE Decision Support Unit. 2014.
20. Collett D. Modelling Survival Data in Medical Research. Chapman and Hall. 2003.
21. Latimer N, Abrams K. NICE DSU Technical Support Document 16: Adjusting Survival Time Estimates In The Presence Of Treatment Switching. NICE Decision Support Unit. 2014.
22. Latimer N, Abrams K, Lambert P, et al. Adjusting Survival Time Estimates to Account for Treatment Switching in Randomized Controlled Trials - an Economic Evaluation Context: Methods, Limitations, and Recommendations Medical Decision Making. 2014;34(3):387-402.
23. Guyot P, Ades A, Beasley M, Lueza B, Pignon J-P, Welton N. Extrapolation of Survival Curves from Cancer Trials Using External Information. Medical Decision Making. 2016.
24. Bagust A, Beale S. Survival Analysis and Extrapolation Modeling of Time-to-Event Clinical Trial Data for Economic Evaluation: An Alternative Approach. Medical Decision Making. 2013.
25. Latimer N. NICE DSU Technical Support Document 14: Survival Analysis For Economic Evaluations Alongside Clinical Trials - Extrapolation With Patient-Level Data NICE Decision Support Unit. 2011.
26. Latimer N. Response to ‘‘Survival Analysis and Extrapolation Modeling of Time-to-Event Clinical Trial Data for Economic Evaluation: An Alternative Approach’’ by Bagust and Beale. Medical Decision Making. 2014;34(3):279-82.
27. Briggs A, Sculpher M, Claxton K. Decision Modelling for Health Economic Evaluation. Oxford University Press. 2006.
28. Broglio K, Berry D. Detecting an Overall Survival Benefit that Is Derived From Progression-Free Survival. Journal of the National Cancer Institute. 2009;101(23).
29. Dejardin D, Lesaffre E, Verbekea G. Joint modeling of progression-free survival and death in advanced cancer clinical trials. Statistics in Medicine. 2010;29:1724-34.
30. Meira-Machado L, Uña-Álvarez Jd, Datta S. Nonparametric estimation of conditional transition probabilities in a non-Markov illness-death model. Computational Statistics. 2015;30:377-97.

31. Putter H, van der Hage J, de Bock G, Elgalta R, van de Velde C. Estimation and Prediction in a Multi-State Modelfor Breast Cancer. Biometrical Journal. 2006;48(3):366-80.
32. Putter H, Fiocco M, Geskus R. Tutorial in biostatistics: Competing risks and multi-state models. Statistics in Medicine. 2007;26:2389–430.
33. Putter H. Special Issue about Competing Risks and Multi-State Models. Journal of Statistical Software. 2011;38(1).
34. Williams C, Lewsey J, Briggs A, Mackay D. Cost-effectiveness analysis in R using a multi-state modelling survival analysis framework: A tutorial. Medical Decision Making. 2016.
35. Asaria M, Walker S, Palmer S, et al. Using electronic health records to predict costs and outcomes in stable coronary artery disease. Heart. 2016;0:1-8.
36. Castelli C, Combescure C, Foucher Y, Daures J. Cost-effectiveness analysis in colorectal cancer using a semi-Markov model. Statistics in Medicine. 2007;26:5557–71.
37. Bongers M, de Ruysscher D, Oberije C, Lambin P, Uyl–de Groot C, Coupe´ V. Multistate Statistical Modeling: A Tool to Build a Lung Cancer Microsimulation Model That Includes Parameter Uncertainty and Patient Heterogeneity. Medical Decision Making. 2016;36:86-100.
38. Williams C, Lewsey J, Mackay D, Briggs A. Estimation of Survival Probabilities for Use in Cost-Effectiveness Analyses A Comparison of a Multi-State Modeling Survival Analysis Approach with Partitioned Survival and Markov Decision-Analytic Modeling. Medical Decision Making. 2016.
39. Meira-Machado L, de Uña-Álvarez J, Cadarso-Suárez C, Andersen P. Multi-state models for the analysis of time-to-event data. Statistical Methods in Medical Research. 2009;18(2):195-222.
40. Meier-Hirmer C, Schumacher M. Multi-state model for studying an intermediate event using time-dependent covariates: application to breast cancer. BMC Medical Research Methodology. 2013;13(80).
41. Hu C, Tsodikov A. Joint modeling approach for semicompeting risks data with missing nonterminal event status. Lifetime Data Analysis. 2014;20:563-83.
42. Jazic I, Schrag D, Sargent D, Haneuse S. Beyond Composite Endpoints Analysis: Semicompeting Risks as an Underutilized Framework for Cancer Research. Journal of the National Cancer Institute. 2016;108(12).
43. Jackson C. Multi-State Models for Panel Data: The msm Package for R. Journal of Statistical Software. 2011;38(8).
44. Welton N, Ades A. Estimation of Markov Chain Transition Probabilities and Rates from Fully and Partially Observed Data: Uncertainty Propagation, Evidence Synthesis, and Model Calibration. Medical Decision Making. 2005;25:633-45.
45. Titman A, Sharples L. Model diagnostics for multi-state models. Statistical Methods in Medical Research. 2009;00:1-31.
46. Ades A, Welton N, Caldwell D, Price M, Goubar A, Lu G. Multiparameter evidence synthesis in epidemiology and medical decision-making. Journal of Health Services Research & Policy. 2008;13(3):12-22.

47. García-Albéniz X, Maurel J, Hernán M. Why post progression survival and post relapse survival are not appropriate measures of efficacy in cancer randomized clinical trials. International Journal of Cancer. 2015;136(10):2444-7.
48. Faria R, Hernandez Alava M, Manca A, Wailoo A. NICE DSU technical support document 17: the use of observational data to inform estimates of treatment effectiveness for technology appraisal: methods for comparative individual patient data. Sheffield: NICE Decision Support Unit. 2015.
49. Kleinbaum D, Klein M. Survival analysis: a self-learning text: Springer Science & Business Media; 2006.
50. Wang W. Nonparametric estimation of successive duration times under dependent censoring. Biometrika. 1998;85(3):561-72.
51. Hawkins N, Sculpher M, Epstein D. Cost-Effectiveness Analysis of Treatments for Chronic Disease: Using *R* to Incorporate Time Dependency of Treatment Response. Medical Decision Making. 2005;25:511–9.
52. Nivolumab for previously treated locally advanced or metastatic squamous non-small cell lung cancer [ID811]. Liverpool Reviews and Implementation Group. STA Report.
53. Goeree R, Villeneuve J, Goeree J, et al. Economic evaluation of nivolumab for the treatment of second-line advanced squamous NSCLC in Canada: a comparison of modeling approaches to estimate and extrapolate survival outcomes. Journal of Medical Economics. 2016.
54. Coyle D, Coyle K. The inherent bias from using partitioned survival models in economic evaluation. Value in Health. 2014;17(3):A194.
55. Briggs A, Baker T, Gilloteau I, Orsini L, Wagner S, Paly V. Partitioned survival versus state transition modeling in oncology: A case study with nivolumab in advanced melanoma. ISPOR 18th Annual European Congress, Milan, Italy. 2015.
56. Kaltenthaler E, Tappenden P, Paisley S, Squires H. NICE DSU technical support document 13: identifying and reviewing evidence to inform the conceptualisation and population of cost-effectiveness models. Sheffield: Decision Support Unit, ScHARR, University of Sheffield. 2011.
57. Roberts M, Russell L, Paltiel D, Chambers M, McEwan P, Krahn M. Conceptualizing a model a report of the ISPOR-SMDM modeling good research practices task force–2. Medical Decision Making. 2012;32(5):678-89.

APPENDICES

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

TA	Title	Date Issued	Available from
TA284	Bevacizumab in combination with paclitaxel and carboplatin for first-line treatment of advanced ovarian cancer	May-13	https://www.nice.org.uk/guidance/ta284
TA285	Bevacizumab in combination with gemcitabine and carboplatin for treating the first recurrence of platinum-sensitive advanced ovarian cancer	May-13	https://www.nice.org.uk/guidance/ta285
TA295	Everolimus in combination with exemestane for treating advanced HER2-negative hormone-receptor-positive breast cancer after endocrine therapy	Aug-13	https://www.nice.org.uk/guidance/ta295
TA296	Crizotinib for previously treated non-small-cell lung cancer associated with an anaplastic lymphoma kinase fusion gene	Sep-13	https://www.nice.org.uk/guidance/ta296
TA299	Bosutinib for previously treated chronic myeloid leukaemia	Nov-13	https://www.nice.org.uk/guidance/ta299
TA306	Pixantrone monotherapy for treating multiply relapsed or refractory aggressive non-Hodgkin's B‑cell lymphoma	Feb-14	https://www.nice.org.uk/guidance/ta306
TA307	Aflibercept in combination with irinotecan and fluorouracil-based therapy for treating metastatic colorectal cancer that has progressed following prior oxaliplatin-based chemotherapy [TA307]	Mar-14	https://www.nice.org.uk/guidance/ta307
TA309	Pemetrexed maintenance treatment following induction therapy with pemetrexed and cisplatin for non-squamous non-small-cell lung cancer	Apr-14	https://www.nice.org.uk/guidance/ta309
TA310	Afatinib for treating epidermal growth factor receptor mutation-positive locally advanced or metastatic non-small-cell lung cancer	Apr-14	https://www.nice.org.uk/guidance/ta310
TA311	Bortezomib for induction therapy in multiple myeloma before high-dose chemotherapy and autologous stem cell transplantation	Apr-14	https://www.nice.org.uk/guidance/ta311
TA316	Enzalutamide for metastatic hormone‑relapsed prostate cancer previously treated with a docetaxel‑containing regimen	Jul-14	https://www.nice.org.uk/guidance/ta316
TA319	Ipilimumab for previously untreated advanced (unresectable or metastatic) melanoma	Jul-14	https://www.nice.org.uk/guidance/ta319
TA321	Dabrafenib for treating unresectable or metastatic BRAF V600 mutation‑positive melanoma	Oct-14	https://www.nice.org.uk/guidance/ta321
TA326	Imatinib for the adjuvant treatment of gastrointestinal stromal tumours (review of NICE technology appraisal guidance 196)	Nov-14	https://www.nice.org.uk/guidance/ta326
TA333	Axitinib for treating advanced renal cell carcinoma after failure of prior systemic treatment	Feb-15	https://www.nice.org.uk/guidance/ta333
TA338	Pomalidomide for relapsed and refractory multiple	Mar-15	https://www.nice.org.uk/guidance/ta338

TA	Title	Date Issued	Available from
	myeloma previously treated with lenalidomide and bortezomib		
TA343	Obinutuzumab in combination with chlorambucil for untreated chronic lymphocytic leukaemia	Jun-15	https://www.nice.org.uk/guidance/ta343
TA344	Ofatumumab in combination with chlorambucil or bendamustine for untreated chronic lymphocytic leukaemia	Jun-15	https://www.nice.org.uk/guidance/ta344
TA347	Nintedanib for previously treated locally advanced, metastatic, or locally recurrent non‑small‑cell lung cancer	Jul-15	https://www.nice.org.uk/guidance/ta347
TA357	Pembrolizumab for treating advanced melanoma after disease progression with ipilimumab	Nov-15	https://www.nice.org.uk/guidance/ta357
TA359	Idelalisib for treating chronic lymphocytic leukaemia	Oct-15	https://www.nice.org.uk/guidance/ta359
TA360	Paclitaxel as albumin-bound nanoparticles in combination with gemcitabine for previously untreated metastatic pancreatic cancer	Oct-15	https://www.nice.org.uk/guidance/ta360
TA366	Pembrolizumab for advanced melanoma not previously treated with ipilimumab	Nov-15	https://www.nice.org.uk/guidance/ta366
TA370	Bortezomib for previously untreated mantle cell lymphoma		https://www.nice.org.uk/guidance/ta370
TA371	Breast cancer (HER2 positive, unresectable) - trastuzumab emtansine (after trastuzumab & taxane)	Dec-15	https://www.nice.org.uk/guidance/ta371
TA374	Erlotinib and gefitinib for treating non-small-cell lung cancer that has progressed after prior chemotherapy	Dec-15	https://www.nice.org.uk/guidance/ta374
TA377	Enzalutamide for treating metastatic hormone-relapsed prostate cancer before chemotherapy is indicated	Jan-16	https://www.nice.org.uk/guidance/ta377
TA378	Ramucirumab for treating advanced gastric cancer or gastro–oesophageal junction adenocarcinoma previously treated with chemotherapy	Jan-16	https://www.nice.org.uk/guidance/ta378
TA381	Olaparib for maintenance treatment of relapsed, platinum-sensitive, BRCA mutation-positive ovarian, fallopian tube and peritoneal cancer after response to second-line or subsequent platinum-based chemotherapy	Jan-16	https://www.nice.org.uk/guidance/ta381
TA384	Melanoma (advanced, unresectable, metastatic) - nivolumab	Feb-16	https://www.nice.org.uk/guidance/ta384

A.2 Additional details on TAs using the partitioned survival approach

TA no. (STA unless stated otherwise)	TA285	TA295	TA296	TA299	TA306
Indication	Advanced ovarian cancer	Advanced or metastatic breast cancer	Locally advanced or metastatic non-small-cell lung cancer	Chronic myeloid leukaemia (chronic, accelerated and blast phase)	Advanced stage non-Hodgkin's B‑cell lymphoma
Method correctly described[1]	Yes	No – described as Markov model	Yes	No – described as semi-Markov model	No – described as semi-Markov model
Justification provided[1] *****	No – only use in previous appraisals as justification	Yes – PartSA allows time-dependency of transitions	No – only use in previous appraisals as justification	No – only use in previous appraisals as justification	No – only use in previous appraisals as justification
Survival endpoints modelled	2 (PFS, OS)	2 (PFS, OS)	2 (PFS, OS)	2 (TTD, OS)	3 (TTD, PFS, OS)
Treatment effects on all transitions?	Yes	Yes	Yes	Yes	Yes
External data	No external data	Treatment effects - MTC + naïve indirect comparison	Treatment effects - MTC	Treatment effects - naïve indirect comparison	No external data
Adjustment to endpoint data	No	Avoid curves crossing – PFS benefit translated to OS benefit based on published ratio	Treatment switching	Surrogacy – OS based on treatment response for a subgroup of population	No
PSA	Yes - ERG commented that for PSA simulations where PFS was estimated to be greater than OS, the company assumed OS equal to PFS	Yes - Independent simulations	Yes - Independent simulations	Yes – issue of OS and PFS correlation being ignored is mentioned	Yes - Independent simulations
Concerns raised by ERG / Committee	Uncertain OS extrapolation – OS data used not from latest data cut-off	OS adjustment - ERG criticised and Committee disregarded OS adjustment	Uncertain OS extrapolation - Due to immaturity and treatment switching adjustment	Uncertain OS extrapolation / TE duration - Critique of OS methods, validity of surrogate relationship and post-treatment benefit	Clinical data robustness – such that introduced uncertainty in validity and robustness of OS results

TA no. (STA unless stated otherwise)	TA307	TA309	TA310	TA316	TA321
Indication	Metastatic colorectal cancer	Non-small-cell lung cancer maintenance treatment	Locally advanced or metastatic non-small-cell lung cancer	Metastatic prostate cancer	Unresectable or metastatic BRAF V600 mutation-positive melanoma
Method correctly described[1]	No – described as Markov model	No – described as Markov model	Yes	No – described as Markov model	Yes
Justification provided[1] *****	Yes - structure aligned with endpoints of pivotal trial	No – only use in previous appraisals as justification	No – only use in previous appraisals as justification	No – only use in previous appraisals as justification	Yes - Reflects trial endpoints and facilitates incorporation of external data
Survival endpoints modelled	3 (TTD, PFS, OS)	2 (PFS, OS)	2 (PFS, OS)	2 (TTD, OS)	2 (PFS, OS)
Treatment effects on all transitions?	Yes	Yes	Yes	Yes	Yes
External data	No external data	No external data	Treatment effects - MTC	Treatment effects - indirect comparison	Treatment effects + Extrapolation: indirect comparison + registry data + general population mortality
Adjustment to endpoint data	No	No	No	No	Treatment switching
PSA	Yes - Independent simulations	Yes - Independent simulations	Yes - Independent simulations	Yes - Independent simulations	Yes
Concerns raised by ERG / Committee	Uncertain OS extrapolation / TE duration - uncertain assumption that treatment benefit continues beyond trial period	TE duration - no evidence to support a post-progression benefit were provided.	Uncertain OS extrapolation – concern for assumption of proportional hazards	OS extrapolation – issue of constant vs. time-dependent OS HR versus BSC	OS extrapolation – the most recent data cut-off was not used

TA no. (STA unless stated otherwise)	TA333	TA338	TA344	TA347	TA357
Indication	Advanced renal cell carcinoma	Advanced multiple myeloma	Chronic lymphocytic leukaemia, all stages	Locally advanced, metastatic, or locally recurrent non-small-cell lung cancer	Advanced melanoma
Method correctly described[1]	Yes	Yes	No – described as semi-Markov model	Yes	Yes
Justification provided[1] *****	No – only use in previous appraisals as justification	No – only use in previous appraisals as justification	No – only use in previous appraisals as justification	No – only use in previous appraisals as justification	Yes - in line with clinical endpoints in pivotal trial
Survival endpoints modelled	2 (PFS, OS)	3 (TTD, PFS, OS)	2 (PFS, OS)	2 (PFS, OS)	2 (PFS, OS)
Treatment effects on all transitions?	Yes	Yes	Yes	Yes	Yes
External data	Treatment effects - indirect comparison or simulated treatment comparison (STC) for different patient subgroups	Treatment effects - observational study for comparator therapy	Treatment effects + Extrapolation – published study for comparator therapy	Treatment effects - MTC	Treatment effects + Extrapolation - registry data and pooled long-term survival curves inform OS in extrapolation period / HR for comparator therapy
Adjustment to endpoint data	No	Treatment switching	No	No	Treatment switching
PSA	Yes - Independent simulations	Yes - Independent simulations	Yes – company mentioned that in some PSA iterations OS crossed over PFS	Yes - Independent simulations	Yes - Independent simulations
Concerns raised by ERG / Committee	Indirect comparison – STC method is likely to bias OS and PFS results for BSC	Uncertain OS extrapolation - Due to small observational study used to inform comparator	Uncertain OS extrapolation / TE duration – Survival after progression unknown and company provides no evidence for assumption that OS HR=1 after trial follow-up	Uncertain OS extrapolation - Limited trial data / Committee suggested use of registry data	Uncertain OS extrapolation – Critique on use of published pooled long-term OS data and not trial OS data

TA no. (STA unless stated otherwise)	TA359	TA360	TA366	TA371	TA374 (MTA)
Indication	Chronic lymphocytic leukaemia, all stages	Previously untreated metastatic pancreatic cancer	Advanced melanoma	Locally advanced or metastatic breast cancer	Locally advanced or metastatic non-small-cell lung cancer
Method correctly described[1]	No – described as Markov model	Yes	Yes	Yes	NR**– no method stated
Justification provided[1] *****	No – only use in previous appraisals as justification	No – only use in previous appraisals as justification	Yes – approach aligned with clinical endpoints assessed in pivotal trial	No – only use in previous appraisals mentioned as justification	Yes – PartSA used due to data availability; PFS and OS routinely reported
Survival endpoints modelled	3 (TTD, PFS, OS)	3 (TTD, PFS, OS)	2 (PFS, OS)	2 (PFS, OS)	2 (PFS, OS)
Treatment effects on all transitions?	Yes	Yes	Yes – long-term OS assumed equal across treatments when registry data applied	Yes	Yes
External data	Treatment effects - median survival estimates calibrated in pivotal trial to provide estimates for additional comparators	Treatment effects - MTC	Treatment effects+ Extrapolation - Published pooled long-term survival curve + registry data	Treatment effects - MTC	No external data
Adjustment to endpoint data	Treatment switching	No	No	No	No
PSA	Yes - Independent simulations	Yes - Independent simulations	Yes - Independent simulations	Yes - Independent simulations – mention of lack of accounting for PFS and OS correlation	Yes - Independent simulations – issue of OS and PFS correlation mentioned
Concerns raised by ERG / Committee	TE duration - Treatment benefit assumed to continue after treatment discontinuation and trial follow-up, despite trial being terminated early for benefit	Uncertain OS extrapolation - ERG used data only from the period towards the end of the KM survival curve to extrapolate	Uncertain OS extrapolation - Limited trial follow-up (12m)	-	Evidence availability - Not all patient subgroups assessed

TA no. (STA unless stated otherwise)	TA377	TA378
Indication	Metastatic prostate cancer	Advanced gastric cancer
Method correctly described[1]	No – described as Markov model	Yes
Justification provided[1] *****	No – only use in previous appraisals as justification	No – only use in previous appraisals as justification
Survival endpoints modelled	2 (TTD, OS)	2 (PFS, OS)
Treatment effects on all transitions?	Yes	Yes
External data	Treatment effects - naïve indirect comparison	Treatment effects - NMA
Adjustment to endpoint data	Treatment switching	No
PSA	Yes - Independent simulations	Yes - Independent simulations
Concerns raised by ERG / Committee	Uncertain OS extrapolation – Due to immature trial data	-

TA - technology appraisal; STA – single technology appraisal; MTA – multiple technology appraisal; ERG – Evidence Review Group; PSA – probabilistic sensitivity analysis; PartSA - partitioned survival analysis; OS – overall survival; PFS - progression-free survival; TTD - time to treatment discontinuation; TE - treatment effect; NR – not reported; HR – hazard ratio; NMA – network meta-analysis; MTC – mixed treatment comparison; KM – Kaplan-Meier; BSC - Best supportive care; STC – simulated treatment comparison;

[1] As per company submission

* Denotes any justification provided beyond precedent appraisals ** Not reported

A.3 Summary of NICE cancer TAs using non-Part-SA approaches

TA no. (STA unless stated otherwise)	Indication	Model type	Justification Provided	States	Treatment effects applied to endpoints	External data	Concerns raised by ERG/committee
TA284	Advanced ovarian cancer	Markov	Treatment switching*	PF PP Death	PFS**	No	No attempt to adjust for treatment switching or use observed OS
TA311	Multiple myeloma, all stages, adjuvant treatment	Decision tree and PartSA	Immature PFS and OS so model uses response as a surrogate	CR-PF PR-PF NR-PF CR-PP PR-PP NR-PP	Response	External trial for long-term OS	Assumption that response is the sole determinant of OS Use of response rather than other surrogates
TA319	Advanced melanoma	semi- Markov	To reflect subsequent lines of treatment	PP FST SST TST PC (Palliative care) Death	PFS (proxy for change of therapy) OS Second-line OS	External trials to inform effectiveness of comparator therapies. Registry data for long-term OS (for all treatments). Previous TA to inform OS adjustment for subsequent lines of therapy.	OS assumptions applied to subsequent therapy lines Relative efficacy assumptions
TA326	Gastrointestinal stromal tumours, adjuvant treatment	Markov	Immature OS	RF PR Death	RFS**	External trials for PPS	Uncertain OS extrapolation Methods for adjustment for treatment switching

TA343	Chronic lymphocytic leukaemia, all stages	Markov	Immature OS	PF PP Death	PFS PrePS	External trial for PPS	-
TA370	Mantle cell lymphoma, stage II, III or IV	semi- Markov	Immature OS Implausible OS predictions	PF PP Dead	TTP PrePS	No (efficacy assumptions for additional comparators)	Uncertain OS Separate modelling prePS, PPS Treatment independent PPS
TA381	Advanced ovarian cancer	semi- Markov	Immature OS Represent maintenance cancer treatments	PF FST SST Death	PFS (time to event, proportion death events)	No	Exclusion of direct PFS and OS trial data Questionable assumptions regarding mortality for subsequent therapies
TA384	Advanced melanoma	semi- Markov	Immature OS	PF PP Death	IPD available: TTP PrePS PPS External data: PFS OS	Treatment effects + External trial and registry for extrapolation	Uncertain OS extrapolation

* Company stated that due to large proportion of patients switching from placebo to active treatment in the trial, they assumed an equal rate of death post-progression for both treatments.

** These studies modelled time to first endpoint as a composite of progression/recurrence and death. It is not clear how the proportion of these events that were deaths or progressions was determined in TA284 or TA343, or whether a difference in this proportion across treatments was modelled.

TA - technology appraisal; STA – single technology appraisal; MTA – multiple technology appraisal; ERG – Evidence Review Group; PF – progression-free; PP – post-progression; PC - palliative care; TTP – time to progression; PrePS – pre-progression survival; PPS – post-progression survival; CR – complete response; PR – partial response; NR – no response; OS – overall survival; PFS - progression-free survival; RFS – recurrence-free survival; TTD - time to treatment discontinuation; IPD – individual patient data; FST – first subsequent therapy; SST – second subsequent therapy; TST – third subsequent therapy.