Making better use of evidence to reflect heterogeneity and uncertainty in survival predictions within cost-effectiveness models

Beth Woods, BSc, MSc

PhD by Publication

University of York

Health Sciences

December 2023

Abstract

The application of survival analysis to estimate the rate at which disease or clinical events occur within cost-effectiveness models has become increasingly sophisticated. Relatively little attention has been paid to how survival analysis methods can be used to reflect underlying disease and clinical processes. This thesis uses three case studies to develop and demonstrate approaches that more explicitly link decision-analytic model structures and accompanying survival analyses to the underlying disease and treatment processes of interest. This allows for more comprehensive use of evidence, and explicit assessment of the effects of heterogeneity and uncertainty. The developed approaches allow disease and treatment mechanisms driving heterogeneity in event risk and relative treatment effects to be reflected within cost-effectiveness analyses. This allows cost-effectiveness results to robustly reflect differences between patients and inform transparent optimised recommendations. The implementation of different decision modelling approaches across the case studies reveals the importance of more explicitly linking the model structure and survival analyses. Partitioned survival models, by directly modelling overall survival, disconnect overall survival from other modelled disease and treatment processes. This can limit the potential to use evidence on intermediate outcomes to inform overall survival extrapolations, and limit exploration of how uncertainty in the extrapolation period impacts on decision uncertainty. State transition models underpinned by multi-state survival analysis can allow a fuller use of evidence and exploration of uncertainty, but can introduce practical challenges and technical uncertainties. The choice of when to implement each approach requires a model conceptualisation process that considers the anticipated importance of heterogeneity in survival outcomes, the nature of direct and external survival evidence, and the key areas of uncertainty. The importance of this work has been recognised by a broad range of decision makers and has directly informed NICE Technology Appraisal recommendations; clinical guidelines and methods guidance.

Author's declaration

The five papers that form the core of this thesis are listed below. Whilst I am not the first author on all of the papers, I have significantly contributed to a substantial part of the work and led on the design, development and conduct of the analyses relevant to the thesis topic. My contribution to each paper is described after the citation of the respective paper. I declare that this thesis is a presentation of original work and that the integrative chapter that binds these papers into the thesis is solely my own work. This work has not previously been presented for an award at this, or any other, University. All sources are acknowledged as References.

Paper 1: Woods B, Hawkins N, Dunlop W, O'Toole A, Bramham-Jones S. Bendamustine versus chlorambucil for the first-line treatment of chronic lymphocytic leukemia in England and Wales: a cost-utility analysis. Value in health 2012; 15(5):759-70.
DOI: https://doi.org/10.1016/j.jval.2012.03.1389
[Contribution of the candidate: Design and programming of all statistical analyses and cost-effectiveness modelling, led preparation of the manuscript and subsequent revisions]

Paper 2: Woods, B., Hawkins, N., Mealing, S., Sutton, A., Abraham, W.T., Beshai, J.F., Klein, H., Sculpher, M., Plummer, C.J. and Cowie, M.R. Individual patient data network meta-analysis of mortality effects of implantable cardiac devices. Heart 2015; 101(22):1800-6.
DOI: http://dx.doi.org/10.1136/heartjnl-2015-307634
[Contribution of the candidate: Designed and programmed all statistical analysis including data preparation and undertaking network meta-analysis, led preparation of the manuscript and subsequent revisions]

Paper 3: Mealing, S., Woods, B., Hawkins, N., Cowie, M.R., Plummer, C.J., Abraham, W.T., Beshai, J.F., Klein, H., and Sculpher, M. Cost-effectiveness of implantable cardiac devices in patients with systolic heart failure. Heart 2016; 102(21):1742-9.
DOI: http://dx.doi.org/10.1136/heartjnl-2015-308883
[Contribution of the candidate: Designed and programmed all survival analysis components of the statistical analysis, collaborated on the design of the other endpoint analyses and cost-effectiveness model as part of the research team, contribution as part of the research team to preparation of the manuscript and subsequent revisions]

Paper 4: Woods, B. S., Eleftherios, S., Palmer, S., Latimer, N., and Soares, M. Partitioned Survival and State Transition Models for Healthcare Decision Making in Oncology: Where Are We Now? Value in Health 2020; 23(12): 1613-21.
DOI: https://doi.org/10.1016/j.jval.2020.08.2094
[Contribution of the candidate: development of the description and critique of modelling approaches and recommendations, conducted reviews of NICE appraisals and published studies, led preparation of the manuscript and subsequent revisions]

** Note that a related piece of work was published as a NICE Technical Support Document: Woods B, Sideris E, Palmer S, Latimer N, Soares M. Partitioned survival analysis for decision modelling in health care: a critical review. NICE DSU Technical Support Document. 2017;19:72.*

Paper 5: Woods, B.S., Sideris E., Sydes M.R., Gannon, M.R., Parmar, M.K.B., Alzouebi, M. et al. Addition of Docetaxel to First-line Long-term Hormone Therapy in Prostate Cancer (STAMPEDE): Modelling to Estimate Long-term Survival, Quality-adjusted Survival, and Cost-effectiveness. European Urology Oncology 2018; 1(6):449-58.
DOI: https://doi.org/10.1016/j.euo.2018.06.004
[Contribution of the candidate: I was promoted to the position of Senior Research Fellow in 2017. As the senior analyst for this project I led on the design and specification of the statistical and cost-effectiveness analysis, supervised ES in specific programming tasks, developed programming for numerous complex aspects of the statistical analysis and cost-effectiveness model, led preparation of the manuscript and subsequent revisions]

Acknowledgements

I would like to thank my supervisors Steve Palmer and Susan Griffin for their continued support, encouragement and many excellent insights during the development of this integrative chapter. I would also like to thank all of those who co-authored or contributed in other ways to the underpinning research papers.

Contents

1. Introduction

Central features of cost-effectiveness analysis to support resource allocation decisions

Cost-effectiveness analysis has become an important consideration when assessing whether new and existing health care interventions should be used in different contexts. The core assessment being made when assessing whether an intervention is cost-effective is whether the health benefits attributable to the use of the intervention outweigh the health opportunity costs incurred as resources used to fund the intervention are not available to serve other health care priorities (Drummond *et al.*, 2015). This assessment of the overall health implications of an intervention, including health opportunity costs, can be summarised quantitively as net health effects, where positive net health effects indicate that an intervention is expected to improve population health accounting for opportunity costs. In order to make this assessment it is necessary to estimate the comparative health benefits of the intervention and alternatives, their resource use and cost implications, and how much health those resources could deliver elsewhere in the health system (a measure of health opportunity cost) (Drummond *et al.*, 2015).

Decision analysis provides a quantitative framework for integrating evidence to assess the implications of different courses of action. This in turn provides an explicit basis for reflecting the trade-offs associated with different choices. In the context of health care decision making, these trade-offs relate to the different health benefits, health risks and opportunity costs associated with different interventions. Decision models are often used to support assessments of cost-effectiveness as they allow comparison of all relevant policy options, extrapolation across an appropriate time horizon, reflection of heterogeneity, inclusion of a wide range of relevant evidence and exploration of a range of uncertainties (Briggs, Sculpher and Claxton, 2006; Drummond *et al.*, 2015). A number of features can be considered central to ensure decision models are developed in a way that appropriately supports health care resource allocation decisions.

Reflecting relevant disease and clinical processes

Decision models are typically structured to describe key disease or clinical processes and the effect of interventions on these processes. This is particularly important when the model is used to extrapolate beyond the observed data from which treatment efficacy has been estimated (which may take the form of a trial, observational study, or, routinely collected healthcare data) in order to estimate long-term outcomes and costs (Latimer and Adler, 2022). In some instances, the structural links (i.e. mathematical relationships) between disease or clinical events, and the evidence

quantifying these links, will provide the *only* mechanism for generating key model outputs. For example, in hepatitis C, decision models have been used to predict long-term outcomes by combining clinical trial estimates of short-term sustained virologic response (SVR) and broader evidence relating SVR to liver damage and mortality (Faria *et al.*, 2016). In other cases, the evidence from which treatment efficacy has been estimated may include all relevant outcomes, but there may remain a need to extrapolate these data if they are incomplete. An example of this would be a clinical trial collecting evidence on treatment response, progression-free and overall survival in advanced cancer where a proportion of patients remain alive at the end of follow-up. Here, the model structure provides the mechanism for extrapolating patients' experience of health events beyond the period of follow-up (Williams *et al.*, 2016).

In addition, structuring decision models to describe key disease or clinical processes ensures that models can draw upon evidence collected in relation to these processes; that the consequences of these processes for health and resource use are captured; and provides a framework through which different judgements can be explored quantitively within a deliberative process. These judgements may relate to the relevance and quality of different sources of evidence, or the structural links between processes (National Institute for Health and Care Excellence, 2022).

Development of a decision model requires a series of decisions about which disease or clinical processes to reflect and how those should be quantitively represented. This process of model conceptualisation ideally comprises two stages (Kaltenthaler *et al.*, 2011). The first stage involves developing a description of the disease or clinical processes relevant to the impacts of the intervention. The second stage involves specifying a quantitative model design, including anticipated evidence requirements.

Reflecting all relevant evidence

Decision models comprise of a set of parameters, and the structural relationships between those parameters and the final outcomes of interest, such as costs and QALYs. Parameters are quantitative descriptions of disease or clinical processes or their implications for costs and QALYs. They can include the rates at which patients experience health events, the resource consumption associated with specific interventions or health events and the impact of intervention and health events on health-related quality of life (HRQoL). A key principle of decision modelling is that parameters should be estimated in a way that reflects all relevant evidence. Depending on the expected importance of the parameter and the resource and time available for model development, systematic reviews or other systematic approaches are recommended to retrieve relevant evidence (Kaltenthaler *et al.*, 2011).

Where multiple sources of evidence are relevant to a particular parameter or group of parameters, these should be synthesised using appropriate statistical methods (Welton *et al.*, 2012). Significant methodological advances have been made in the development and application of evidence synthesis methods for use within decision models (Welton *et al.*, 2012; Dias *et al.*, 2018). The process of structuring models and assessing the credibility of their parameter inputs and projections should also account for all relevant evidence. Where evidence is non-existent or sparse, the evidence considered relevant may be broadened to include expert opinion or empirical evidence that is indirectly relevant (Nikolaidis *et al.*, 2021) e.g. historical observational evidence, evidence on outcomes not of direct relevance to patients, evidence relating to treatments with a similar mechanism of action (Jackson *et al.*, 2017; Palmer *et al.*, 2023).

Quantifying heterogeneity

Heterogeneity refers to the differences across patients in costs or outcomes that can be associated with observed characteristics. These differences may arise due to differences in underlying prognosis, the relative effects of treatments, costs or HRQoL. Decision models can be used to generate predictions of health benefits and costs in subgroups of patients defined by clinical and/or demographic characteristics. This in turn offers the potential to make different decisions across subgroups where the evidence indicates that an intervention may be cost-effective in some subgroups but not others. Compared to making decisions based on average cost-effectiveness in broader populations, reflecting heterogeneity analytically and in decisions has the potential to generate gains in overall population health (Coyle, Buxton and O'Brien, 2003; Drummond *et al.*, 2015).

Quantifying uncertainty

Any assessment of the health benefits and costs of alternative interventions is uncertain. This uncertainty arises from a number of sources. In the context of decision models, parameter inputs will be associated with uncertainty, as will the structural assumptions within the decision model and any underlying statistical models (Drummond *et al.*, 2015). One considerable source of uncertainty in economic evaluations results from incomplete follow up of study subjects for the final outcomes of interest. For example, in a review of National Institute for Health and Care Excellence (NICE) technology appraisals of cancer drugs 41% were considered by independent academic reviewers or a NICE Committee to have immature overall survival evidence (Tai *et al.*, 2021).

Characterising this uncertainty is required to correctly estimate expected health benefits and costs[1], to determine the robustness of the results to alternative assumptions (parameter and structural) and can also be used to assess the value of conducting further research. The value of research can be assessed quantitatively by using decision models to generate estimates of the value of information (Claxton, 1999). Quantifying the net health implications of investing in research alongside the net health implications of adoption allows a broader set of policy options to be evaluated (Claxton *et al.*, 2012). These include (i) adopt without further research; (ii) approve alongside further research and (iii) delay widespread use until research reports (only in research). The consequences of these policies will depend on the net health benefits of adoption and research but also any opportunity costs associated with early adoption (which may include the reduced likelihood of research going ahead once a technology is adopted, the possibility that decisions cannot be reversed even if new research suggests an intervention is not cost-effective, or sunk costs incurred due to adoption). An explicit framework has been developed to allow these considerations to be assessed quantitatively (Claxton *et al.*, 2012).

The role of survival analysis in decision models

Decision models often characterise conditions using a series of clinical states or events (e.g. mortality, cardiovascular events, cancer progression, relapses from symptom control, treatment discontinuation) (Briggs, Sculpher and Claxton, 2006; Siebert *et al.*, 2012). For many interventions the rate at which patients progress through these health states, and how this varies across interventions, will be one of the key determinants of cost-effectiveness, as this determines life expectancy, health care costs and HRQoL.

Evidence on the occurrence of clinical events is often available as survival data which describe the time until a health-related event occurs. Survival data have particular features and a set of statistical methods called survival analysis are available to analyse these data. The distinct feature of survival data is that patients may not reach the endpoint of interest due to study completion or loss to follow up. In these instances, the time to event is unknown, though the patient still contributes important information, as we know that their survival time is at least as long as the time for which they were observed. This is known as right censoring (Collett, 2003). In the context of clinical studies, survival data are typically described using non- or semi-parametric statistical approaches such as Kaplan

[1] Use of average parameter values (rather than propagating the uncertainty in parameter inputs through the model and then averaging across the resulting model outputs) can result in inaccurate model outputs where there are non-linear relationships between parameter inputs and model outputs.

Meier estimates and Cox proportional hazards models. These methods are often selected as they can be used to assess the statistical significance of differences between interventions without making assumptions about how event rates change over time (more formally, the hazard function) and in the case of the Cox model assume only that hazards are proportional between interventions. However, they have limited utility in the context of decision models where there is a need to estimate absolute differences between interventions, and to extrapolate event rates beyond the period observed in the study.

Extrapolation is necessary where differences in clinical events between interventions are expected to persist beyond the period of study follow up (Latimer and Adler, 2022). This may be due to ongoing effects of treatments on event rates, or ongoing prognostic differences due to differences across treatment in health state occupancy at the end of study follow up.

Parametric survival models assume that survival times follow a specific distribution which allows the hazard and survival function to be expressed as a mathematical function of time. This allows survival and other related quantities to be predicted at any time point including beyond the period of study follow up. Parametric survival models have been used extensively within decision models to describe the evolution of the hazard over time and facilitate extrapolation (Bell Gorrod *et al.*, 2019).

Where a range of comparators are considered relevant, network meta-analysis (NMA) is often used within decision models to inform comparative effectiveness (Drummond *et al.*, 2015). Network meta-analyses of survival outcomes frequently synthesise hazard ratio estimates that were generated by fitting Cox models within individual studies. Hazard ratio estimates from a NMA can be applied to a baseline parametric estimate of the hazard function (Woods *et al.*, 2017).

Over the last 15 years there have been significant advancements in the use of parametric survival models to inform decision models, in recognition that choice of parametric model can be an important driver of cost-effectiveness analysis and reimbursement decision making. This has emphasised the importance of systematically assessing and selecting between alternative parametric models; assessing the credibility of long-term extrapolations using external data; and has examined the role of more complex parametric models (Jackson *et al.*, 2017; Latimer, 2011; Rutherford *et al.*, 2020). Network meta-analysis methods have been developed that allow hazard ratios to vary over time (Freeman *et al.*, 2022). These methods may be relevant where an assumption of proportional hazards across treatment options is not considered appropriate.

Expanding survival analytic approaches to support decision making

The application of survival analysis to estimate the rate at which disease or clinical events occur within decision models has become increasingly sophisticated in recent years with a focus on the development and selection of statistical models (Latimer, 2011; Rutherford *et al.*, 2020). This body of work has focused on modelling single clinical events (or composite events) in the context of a single population with a view to providing plausible long-term extrapolations. The uncertainty surrounding extrapolations has been highlighted as a major source of decision uncertainty in many health technology appraisals (HTAs) (Bell Gorrod *et al.*, 2019). Best practice guidance recommends that uncertainty in the parameters of survival models should be reflected in probabilistic sensitivity analysis, and that structural uncertainties relating to the choice of parametric form and duration of treatment effect be examined via scenario analyses (Latimer, 2011).

The central features of cost-effectiveness analysis to support resource allocation decisions outlined above have received relatively little attention. Despite the statistical models being developed to inform decision models, little consideration has been given to how these statistical models should reflect the interconnected disease and clinical processes that the decision models aim to describe.

Heterogeneity in clinical event rates or the effects of treatment is rarely the focus of methodological developments or applied work (Ward *et al.*, 2021). Though best practice guidance has emphasised the importance of exploring the sensitivity of model results to alternative plausible extrapolations, this has focused on the statistical specification of models that estimate time to an individual endpoint. This has led to a focus on time trends in those endpoints rather than a more comprehensive consideration of available evidence. For example, analyses of overall survival may ignore relevant evidence relating to intermediate endpoints or external datasets. Limited attention has been paid to how choices of decision model structures and accompanying survival analyses can reflect uncertainties in the disease and treatment processes. For example, models that directly extrapolate observed overall survival data rather than predicting overall survival as a function of disease and treatment processes, may be ill equipped to address uncertainties relating to the receipt of subsequent health care interventions as these are not explicitly modelled.

Hence there is a need to develop and apply survival analytic approaches within decision models that reflect all relevant evidence, allow heterogeneities in event risks to be reflected in cost-effectiveness results, and quantify the broad range of uncertainties inherent in model extrapolations. This is important to ensure decision making: is informed by the full range of evidence; can focus intervention access on those patient populations in whom usage is expected to deliver population

health benefits; and, can ensure resource allocation and research decisions reflect extrapolation uncertainties.

Aims, objectives and structure of the thesis

The aim of this thesis is to provide a coherent and original contribution to advancing the methods, application and interpretation of cost-effectiveness analysis informed by survival data. More specifically, the objectives of the thesis are to demonstrates how more explicitly linking model structures and accompanying survival analyses to the underlying disease and treatment processes of interest can be achieved and how this can allow for more comprehensive use of evidence, and assessment of the effects of heterogeneity and uncertainty. This in turn should support better decisions by facilitating the development of models that more appropriately characterise costs and outcomes, and their uncertainty. The remainder of the thesis is structured in three sections followed by a discussion:

- Section 2 develops and applies decision modelling and survival analytic methods to address each of the central features outlined above. This section draws on the case study presented in Paper 1 which examines the cost-effectiveness of bendamustine compared to chlorambucil as a first line treatment for chronic lymphocytic leukaemia (CLL) and quantifies the value of further research to support decision making in this setting. Paper 1 uses semi-Markov modelling informed by interrelated parametric survival analyses. These approaches allow the cost-effectiveness analysis to reflect the effects of patients' depth of response and re-treatment/subsequent treatments on HRQoL and costs; to reflect heterogeneity and to reflect uncertainties in the disease and treatment process. As well as highlighting the contributions of paper 1 this section identifies a series of areas requiring further development which are the focus of sections 3 and 4.

- Section 3 develops and applies a cohesive set of statistical and decision analytic modelling methods to allow a more comprehensive exploration of heterogeneity arising from patient risk factors and the intervention mechanisms of action, using evidence from multiple randomised controlled trials (RCTs). This section draws on the case study presented across Papers 2 and 3 of the use of implantable cardioverter defibrillators (ICD) and cardiac resynchronisation therapy (CRT) for people with ventricular arrhythmias or heart failure. Papers 2 and 3 demonstrate how parametric survival modelling and individual patient data (IPD) NMA can be combined to provide a more comprehensive assessment of heterogeneity. This section also develops approaches for presenting large numbers of subgroup results to decision makers.

- Section 4 appraises alternative methods for using survival analysis within decision models to reflect interconnected disease and treatment processes. This section draws on methodological and empirical work presented in papers 4 and 5. Paper 4 critically appraises Partitioned Survival Modelling (PSM), a frequently used modelling approach in oncology, and assesses the relative merits of an alternative approach - state transition models (STMs) underpinned by multi-state survival analysis. Paper 5 demonstrates the application of an STM and multi-state survival analysis within a cost-effectiveness analysis of docetaxel in prostate cancer patients initiating hormone therapy.

The discussion section summarises the contribution of the thesis and provides guidance on how those tasked with developing cost-effectiveness analyses to support decision making can assess when the more complex approaches presented within this thesis may be appropriate.

2. Developing decision and survival models to address the central features of cost-effectiveness analysis: a case study in chronic lymphocytic leukaemia

Paper 1 examines the cost-effectiveness of bendamustine when compared to chlorambucil as a first line treatment for CLL and quantifies the value of further research to support decision making in this setting. Decision modelling was necessary as the pivotal clinical trial (02CLLIII) provided immature outcome data (i.e. a large proportion of patients had not experienced all clinical events of interest) and very limited evidence on HRQoL or resource use.

This paper departs from typical approaches to decision modelling in oncology which often focus on disease progression and death as the events of interest (Hoyle *et al.*, 2010). A semi-Markov model was developed that tracked depth of response to treatment, progression, and the occurrence and outcomes of further lines of treatment (see Figure 1, Paper 1). This more detailed structure allowed the model to reflect the increased HRQoL and durability associated with deeper remissions, and the cost and HRQoL implications of receipt of re-treatment and subsequent lines of treatment. This was important in the assessment of incremental value as the choice of re-treatment was dependent on first line treatment choice and outcomes.

For the base case analysis, parametric survival models were developed using IPD from the 02CLLIII trial to estimate time to progression in each response group, time from progression to re-treatment

and overall survival. Systematically identified evidence from the literature was used to provide information on outcomes following re-treatment and subsequent lines of treatment. Parametric survival analyses of the overall survival endpoint were used to directly estimate the probability of death, independently of patients experience of other clinical events. This ensured a close correspondence between model predictions of overall survival and the Kaplan Meier estimates of survival for the trial follow-up period (Figure 3, Paper 1). The competing risk nature of progression (or re-treatment) and death events was accounted for by considering death as a censoring event in the time to progression and time from progression to re-treatment parametric survival analyses. The probability of experiencing a non-fatal event in a given model cycle was then calculated as the predicted probability of remaining alive multiplied by the probability of experiencing the event of interest (with each probability calculated from the corresponding parametric model).

Given the population in focus for this evaluation – patients unfit for more aggressive treatment – the analysis examined whether bendamustine was cost-effective across patients with different characteristics considered to be predictive of their frailty. Cost-effectiveness results were generated for three subgroups: people aged 65 years or older; people with a World Health Organisation (WHO) performance status of 1 or higher; and people aged 65 years or older who also had a WHO performance status of 1 or higher. These analyses used subgroup-specific estimates of treatment response and included a variable indicating membership of the relevant subgroup within the parametric survival modelling.

As well as providing a more accurate representation of patients' cost and HRQoL trajectories, the more detailed modelling of the underlying disease and treatment process allowed for a more thorough examination of key parameter and structural uncertainties. This included examination of the nature of the effects of first line treatment choice on patient outcomes (e.g. did bendamustine influence only response probability or also time to progression conditional upon response?), the duration of response that would support re-treatment, and outcomes following subsequent lines of treatment. Estimates of the expected value of perfect information (EVPI) were also presented for different measures of health opportunity cost to inform assessments of whether further research could represent a worthwhile use of health care resources.

The analyses developed provided the basis for decision making in the NICE technology appraisal TA216. An independent review of the model within this process found the model to be high quality, more sophisticated than previous models, and that the modelling of the disease and treatment process better reflected the realities of CLL disease management than previous models (National Institute for Health and Care Excellence, 2011). The NICE committee concurred with this view

though noted that data available to inform subsequent line therapy outcomes were sparse. The analysis supported the positive appraisal decision for bendamustine which was recommended for routine use in the NHS across the eligible population (National Institute for Health and Care Excellence, 2011). An adapted version of the analyses supported the recommendation of bendamustine for routine use in Scotland (Scottish Medicines Consortium, 2011).

Although this paper addresses all the central features outlined in Section 1 of the thesis, the methods used highlighted a number of areas where the use of survival analysis within decision models could be further developed to meet the needs of decision makers.

In all analyses overall survival was independent of the modelled disease and treatment processes. Overall survival data from the pivotal trial was immature with ~ 60% remaining alive at 5 years. The decision model developed therefore had to extrapolate mortality for the majority of patients over many years. Only the information on the mortality outcome informs these extrapolations, despite the availability of near complete response data and relatively mature data on time to progression. This reflected a lack of available methods to build a robust structural link between intermediate outcomes and survival at the time of this work. This suggests that alternative approaches could be developed that better utilise all available evidence to support overall survival extrapolations. The disconnect between the disease and treatment processes and overall survival also reduces the utility of the analyses undertaken to quantify uncertainty. Analyses examining the sensitivity of model results to treatment outcomes and pathways assumed that overall survival was invariant to these parameters, an assumption which lacks face validity. The assumption of independence between overall survival and intermediate endpoints was also made within the probabilistic sensitivity analysis. It is therefore possible that the estimates of EVPI generated from the model may not have accurately characterised the value of further research.

The analysis of heterogeneity bifurcated the population within 02CLLIII using a series of definitions of frailty and assumed that these markers of frailty did not modify the effect of treatment on time to progression (conditional upon response). An analysis that considered all patient profiles of interest (e.g. patients aged <65 with performance status of zero) may have indicated subgroups in which bendamustine was not cost-effective thus allowing for more refined reimbursement recommendations. In addition, a more extensive exploration of the effects of patient characteristics on model parameters may have come to different conclusions about the cost-effectiveness of bendamustine in specific subgroups.

Conclusions

The survival analyses and decision modelling methods employed in Paper 1 provided significant utility to decision makers by more accurately characterising patients trajectory of treatments and outcomes than previous models, exposing this to sensitivity analysis and exploring key areas of heterogeneity. However, the methods used extrapolated overall survival without considering evidence relating to intermediate endpoints, were unable to capture the effects of clinical uncertainties on overall survival (and therefore long-term costs and QALYs), and provided a limited quantification of heterogeneity. This suggests that further development and application of methods is required to ensure resource allocation and research decisions appropriately reflect all available evidence and characterise heterogeneity and uncertainty appropriately. These methods are further developed in the remainder of the thesis.

3. Developing statistical and decision modelling methods to address heterogeneity and synthesise survival evidence

This section of the thesis develops and applies a cohesive set of statistical and decision analytic modelling methods to allow an extensive exploration of heterogeneity. This section draws on the cardiovascular device case study presented in Papers 2 and 3. The analyses were developed to assess the cost-effectiveness of three implantable cardiac devices: ICDs, CRT pacemakers (CRT-Ps) and combination therapy (CRT-D) in patients with heart failure with reduced ejection fraction. The analyses were undertaken to inform NICE technology appraisal TA314 (National Institute for Health and Care Excellence, 2014). Decision modelling was required to simultaneously compare all devices, synthesise multiple sources of evidence, and, quantify the effects of patient characteristics on costs and outcomes. A thorough exploration of heterogeneity within the cost-effectiveness analysis was considered appropriate as patients eligible for the implantable devices were highly heterogeneous in terms of their baseline mortality risk, and, expected survival benefits of device implantation. It was therefore expected that the most effective and cost-effective intervention choice would vary across subgroups.

The decision model developed has two health states (alive and dead), directly models overall survival, and is parameterised using a series of survival analyses and other regression equations to predict: mortality, hospitalisation rates, and, HRQoL. The statistical models were developed using IPD from a network of 13 RCTs including 12,638 patients (see Figure 1, Paper 2). Ideally given the nature of the mortality effects of the devices, with ICDs preventing sudden cardiac death and CRT preventing death due to pump failure, the model would have reflected individual causes of death.

However, this was not feasible due to missing, inconsistent and unreliable data on cause of death. The potential implications of the mechanisms of action of the devices were instead considered when estimating all-cause mortality parameters and in sensitivity analyses.

Mortality outcomes were the single most influential determinant of cost-effectiveness in this setting (National Institute for Health and Care Excellence, 2014). The decision model is used to predict mortality probabilities for medical therapy and each implantable device for subgroups of patients with different profiles of characteristics. The decision model is parameterised using two survival models: a relative treatment effect model and a baseline risk model. Relative treatment effects are estimated using an adjusted IPD NMA which provides relative treatment effects for each device compared to medical therapy conditional upon a series of patient characteristics. The relative treatment effects are applied to an estimate of baseline mortality risk for patients receiving medical therapy. Baseline risk estimates for medical therapy were derived from a parametric survival model which allowed baseline risk to be extrapolated over time and to vary according to patient characteristics. Baseline clinical and demographic characteristics were selected based on data availability and information on their expected importance as baseline risk or relative treatment effect modifiers obtained from clinical risk scores, trial subgroup analyses and clinical advice.

The NMA of overall survival uses a Cox proportional hazards model stratified by trial. The model synthesises time-to-event IPD and generates estimates of parameters that can be combined to estimate relative treatment effects (hazard ratios) for patients with different baseline characteristics. QRS duration, left bundle branch block morphology (LBBB), age and gender were included as treatment effect modifiers in the final model.

NMA was necessary as no trial had compared all the devices of interest, and allowed inclusion of all relevant trials. IPD NMA is considered the 'gold standard' NMA approach in general, and particularly where patient characteristics that differ across trials are expected to modify relative treatment effects (Debray *et al.*, 2018). This is likely to lead to heterogeneity[2] and inconsistency[3] in the network of evidence, which, in turn, will mean any relative treatment effect estimates derived via NMA are likely to be biased and the population to which they pertain unclear. IPD NMA was therefore necessary to produce meaningful estimates of relative treatment effects in any population and allowed treatment effects to be estimated for relevant subgroups. Alternative simpler approaches suffer from a number of drawbacks. Synthesis of subgroup data is challenging due to inconsistencies in trial reporting, reduces precision, and allows only a univariate exploration of heterogeneity (Riley

[2] Variation in the same treatment effect across studies.
[3] Variation between direct and indirect estimates of the same treatment effect.

et al., 2023). Meta-regression using aggregate data is subject to specific sources of bias and lacks statistical power, particularly when multiple patient characteristics are modelled and/or interaction effects are expected to differ across interventions (Debray *et al.*, 2018; Dias *et al.*, 2013).

The specification of the IPD NMA model allowed the impact of patient characteristics on the efficacy of devices to vary by device (i.e. 'unrelated' interaction terms were modelled (Dias *et al.*, 2013)). This reflected the aforementioned different causes of death targeted by the devices. Interaction terms were, somewhat unusually, expected to differ across devices in direction of effect as well as magnitude. For example, men are more likely to experience sudden cardiac death than women. Due to the competing nature of different causes of death, male participants were therefore expected to experience higher ICD effects and lower CRT effects on all-cause mortality.

Baseline risk of mortality for patients receiving medical therapy was estimated using a parametric model fitted to all patients receiving medical therapy across trials. The final model developed included age, gender, baseline New York Heart Association (NYHA) class, ischaemic aetiology, QRS duration, and left ventricular ejection fraction (LVEF) as covariables. Reflecting heterogeneity in baseline risk is important as this influences the absolute mortality risk reduction associated with intervention and therefore cost-effectiveness (Drummond *et al.*, 2015).

Although much attention has been paid to the selection of the functional form and specification of treatment effects within parametric survival models (Latimer, 2011), little guidance has been given regarding the assessment of models including patient characteristics. This was addressed within Paper 3 using two approaches: statistical tests for the proportional hazards assumption were applied for each patient characteristic, and comparisons of parametric model predictions to Kaplan Meier survival estimates were conducted for each quartile of predicted risk[4] (Collett, 2003). This revealed that, for patients with lower mortality risk, extrapolated mortality fell below that expected amongst the general population, which was considered implausible. To address this, the parametric models were extended to include age as a time-dependent variable. This approach allowed observed information about mortality risk in patients who were older at the time of entry in to the trials to inform the long-term extrapolations of younger patients as they aged in the model. Though this approach produced similar predictions within the period of trial follow up, it produced more plausible predictions in the extrapolation period for individuals at lower mortality risk (Association of British Healthcare Industries, 2012).

[4] Predicted risk was calculated using the linear predictor component of the parametric model.

The IPD NMA predicted relative treatment effects on the hazard ratio scale that varied considerably according to the modelled patient characteristics (Paper 2, Table 2). The parametric modelling of mortality in patients receiving medical therapy predicted that baseline mortality risk also varied considerably according to patient characteristics (Paper 3, Appendix 5). Unlike conventional subgroup analyses which present results stratified by a single characteristic, these multivariate analyses allow the mortality parameters within the cost-effectiveness modelling to simultaneously reflect a patient's profile of characteristics. Model predictions were generated for all combinations of those characteristics included within the statistical estimation models, including characteristics not expected to appear within NICE guidance. This reflected concerns that survival models (in particular) are unlikely to be linear in patients' characteristics so using average covariate values for these characteristics may produce inaccurate estimates of costs and effects. This resulted in the model being run for 4,992 subgroups. Reducing the number of subgroups to a manageable number was important to ensure that results could be appropriately deliberated by a NICE committee. This was achieved by only stratifying results by those characteristics which influenced cost-effectiveness (LVEF and ischemic aetiology were not strong determinants of cost-effectiveness and therefore results were not presented stratified by these characteristics) and that the decision maker was expected to consider within their guidance (results were not presented by age and gender as these are protected characteristics) (National Institute for Health and Care Excellence, 2013). Final results were presented for 24 subgroups defined by NYHA class, QRS duration and presence or absence of LBBB.[5]

Previous cost-effectiveness studies that have conducted a detailed analysis of heterogeneity have generally been designed to support recommendations stratified according to a single measure of event risk (Briggs *et al.*, 2007; Davis *et al.*, 2016). This is appropriate where the main source of heterogeneity is a patients' baseline risk of a single/composite event and clinicians have ready access to an appropriate risk prediction algorithm. This approach was not appropriate in the current context where patient characteristics influenced multiple aspects of the model, and to different degrees. For example, NYHA class has a strong effect on baseline mortality risk but was not included in the treatment effect model whereas QRS had a much smaller effect on baseline mortality risk but was a determinant of treatment effect. We therefore developed the approach outlined above in order to condense the number of subgroups presented to decision makers whilst ensuring that the

[5] Within these subgroups, costs and QALYs were calculated as a weighted average across patients with different ages, genders, LVEFs, and ischemic statuses.

results related to patient subgroups that were clinically meaningful and differentiated with respect to cost-effectiveness.

Uncertainty was addressed through scenarios and probabilistic analyses. An important set of scenario analyses related to the duration of the effects of the devices on all-cause mortality, which were expected to reduce over time as patients aged and were more likely to die of other causes. This effect was not explicitly modelled as the model did not consider separate causes of death. Instead, a gradual waning of the treatment effect was applied and sensitivity analyses were conducted on the point at which the treatment effect begins to wane. Probabilistic sensitivity analyses were used to quantify the effects of parameter uncertainty on predicted costs and outcomes. These analyses showed that although accounting for heterogeneity substantively increased parameter uncertainty, this did not necessarily translate to increased decision uncertainty. For example, for a patient who is female, 65 years old, NYHA class II, has ischemic aetiology, QRS ≥150ms, LVEF 20-25%, and LBBB, the magnitude of treatment effects of CRT-D and ICD on mortality estimated from the IPD NMA are highly uncertain. However, there is little uncertainty as to the preferred device from a cost-effectiveness perspective (Appendix 16 in (Association of British Healthcare Industries, 2012)). In other subgroups decision uncertainty was high. The analysis presented could therefore be used to identify subgroups and parameters on which further research is likely to be of value, however quantifying the value of further research was beyond the scope of these papers.

The analyses developed provided the basis for decision making in NICE TA314. The approach taken allowed the majority of RCT evidence to inform key model parameters (95% of patients identified by the systematic review were included in the analysis) and for a detailed exploration of heterogeneity. The evidence used was recognised as a "rich and important" data source by the NICE committee who also noted that the approach taken "allows consideration of population groups based on clinical characteristics that are considered important by clinicians in current clinical practice for making decisions about device implantation" (National Institute for Health and Care Excellence, 2014). The NICE guidance (Figure 1) clearly distinguished recommendations according to those patient characteristics identified as important determinants of cost-effectiveness within the analyses. Compared to previous decision modelling and guidance these recommendations focused device usage in those patients where it was expected to result in net health gains. For example, in an earlier technology appraisal (TA95), CRT-D was compared to CRT-P in a broad population of NYHA III/IV patients using a model informed by a meta-analysis of aggregate data. CRT-D was not found to be cost-effective at conventional cost-effectiveness thresholds (Fox *et al.*, 2007). Conversely the analyses presented in Papers 2 and 3 found that amongst patients with NYHA class III CRT-D was

cost-effective at a cost-effectiveness threshold of £30,000/QALY, whereas it was not cost-effective amongst patients with NYHA class IV.

	NYHA class			
QRS interval	**I**	**II**	**III**	**IV**
<120 milliseconds	**ICD if there is a high risk of sudden cardiac death**			**ICD and CRT not clinically indicated**
120–149 milliseconds without LBBB	ICD	ICD	ICD	CRT-P
120–149 milliseconds with LBBB	ICD	CRT-D	CRT-P or CRT-D	CRT-P
≥150 milliseconds with or without LBBB	CRT-D	CRT-D	CRT-P or CRT-D	CRT-P
LBBB, left bundle branch block; NYHA, New York Heart Association				

Figure 1: NICE guidance for TA314

The IPD NMA has informed international clinical guidelines (Ezekowitz *et al.*, 2017; McDonagh *et al.*, 2021; Ponikowski *et al.*, 2016). These guidelines had previously relied on informal qualitative synthesis of subgroup data and trial inclusion criteria in order to assess heterogeneity. The IPD NMA has also been used to inform the treatment effect parameters in published cost-effectiveness analyses in Germany and the US (Hadwiger *et al.*, 2021; Shah *et al.*, 2020).

Conclusions

Papers 2 and 3 develop and apply methods for reflecting heterogeneity in survival endpoints within a decision model. The analyses provided an appropriate vehicle for synthesising all relevant evidence, quantifying heterogeneity in a way that reflects the influence of patient characteristics on both mortality risk and response to device therapy, extrapolating mortality risk over patients' lifetimes and characterising uncertainty in costs and effects. Despite the large number of potential subgroups of interest, results were aggregated to the level considered most relevant to decision makers, allowing the analysis to inform clear NICE recommendations.

4. Structuring decision models and survival analyses with multiple survival endpoints

This section of the thesis explores the implications of different decision model structures (and the underpinning survival approaches) for the appropriateness of model extrapolations, and the ability of decision models to quantify uncertainty and heterogeneity. The CLL case study presented in Section 2 used a detailed semi-Markov process to reflect disease and treatment processes but estimated overall survival directly and independently of these processes i.e. using a PSM approach. This meant that overall survival extrapolations were not informed by the evidence available on intermediate endpoints where the evidence was more mature (>80% of patients had experienced a progression-free survival event whereas only ~40% had died), and that the sensitivity analyses did not reflect the effects of clinical uncertainties on overall survival. PSM has been used extensively in HTAs and published cost-effectiveness analyses (Bullement, Cranmer and Shields, 2019), though no guidance was available to determine its appropriateness as a modelling approach to inform policy decisions. Paper 4 sought to address this gap by critically appraising PSM and providing recommendations on PSM implementation and reporting. A key recommendation of Paper 4 is that methods for conducting STM informed by multi-state survival modelling (MSM) should be further developed. Paper 5 presents a case study in prostate cancer demonstrating the feasibility of applying these methods, and this thesis chapter reflects on the advantages and disadvantages of using this approach to support resource allocation decisions.

Appraising alternative modelling approaches

Some findings presented in Paper 4 were originally developed as part of a NICE Technical Support Document (TSD 19) (Woods *et al.*, 2017). TSDs review the state of the art on a topic, and make recommendations on the implementation of methods and reporting standards. Paper 4 summarises findings from TSD 19, reviews more recent literature comparing PSMs and STMs, and reviews more recent NICE appraisals to understand current practice. This section summarises the key learnings from TSD 19 and Paper 4.

PSMs directly use survival analyses of commonly reported time-to-event endpoints (e.g. in oncology this is often the progression-free survival and overall survival endpoints) to derive state membership. The survival curves that inform estimates of state membership are typically modelled completely independently. Application of PSM therefore models overall survival as independent of other clinical events, including those explicitly included within the model structure such as disease progression. This contrasts with STMs where overall survival is predicted using information on the

rate at which patients progress through different health states (or experience events) and the mortality risks associated with these states or events. During the period of study follow up the approaches are expected to produce similar results as relationships between endpoints are reflected within the available data. However, in the extrapolation period, PSM and STM use different sets of information and assumptions, and are, therefore, expected to produce different extrapolations. Recent empirical comparisons of the two methods have confirmed that their extrapolations of overall survival can differ markedly (Paper 4, Table 2).

PSM offers several practical advantages. As the survival curves used within PSM often map directly to clinical trial endpoints, PSMs can be developed without access to IPD and their predictions typically closely follow the study Kaplan Meier survival estimates for the within-study period. However, the lack of biological or clinical structure underpinning PSMs limits the extent to which sensitivity analyses can quantify clinical uncertainties in the extrapolation period, as demonstrated in Section 2.

Paper 4 and TSD19 recommend that, regardless of modelling approach, cost-effectiveness analyses should include tabulations showing the health states in which life year (and QALY) differences between interventions occur, and the extent to which they accrue in the observed vs. extrapolation periods. Survival curves for individual clinical events for each treatment should be produced to support an assessment of whether the degree of benefit in the extrapolation period is clinically plausible and where sensitivity analyses should focus. STMs are recognised as an important vehicle for exploring clinical uncertainties in the extrapolation period. For example, the effects of interventions beyond the point of disease progression are often a significant source of uncertainty which could be explored by varying parameters describing post-progression survival within an STM. The recommendations in Paper 4 do not suggest that STM replace PSM. Appropriate development of STMs requires use of MSM which involves building survival models for each individual transition within the STM. At the time of developing Paper 4 there were considerable uncertainties about the practicalities and appropriate methodologies for applying MSM in the context of decision models, and the available empirical literature indicated that choices made when developing MSMs in this context could strongly influence cost-effectiveness and therefore reimbursement decisions.

The recommendations made within TSD 19 and Paper 4 are now widely cited in submissions to NICE and have informed international HTA guidelines on extrapolation and model structure (Haute Autorité de Santé, 2020; Health Information and Quality Authority, 2020; Neyt *et al.*, 2020). They have also encouraged further methodological work (Majer *et al.*, 2022). Despite the growing interest

in STMs informed by MSM, applications of the approach have taken the form of proof-of-concept studies (Bongers *et al.*, 2016; Castelli *et al.*, 2007; Williams *et al.*, 2016). These studies have focused on statistical estimation rather than the ability of the approach to support decision makers in understanding heterogeneity and uncertainty. The case study in the next section shows the feasibility of applying STM informed by MSM prospectively with the objective of informing resource allocation decisions.

Using state transition models and multi-state survival analysis to inform assessments of cost-effectiveness

Paper 5 shows the feasibility and practicality of applying STM informed by MSM prospectively as the primary modelling approach for a major RCT (the STAMPEDE trial (James *et al.*, 2016)). A decision model was developed using evidence from STAMPEDE and the literature to assess the cost-effectiveness of adding docetaxel to standard of care in men with high risk prostate cancer starting hormone therapy. Decision modelling was necessary to extrapolate beyond the observed data (approximately half of the patients were alive at the point of the primary analysis); and explore heterogeneity in the target population. STAMPEDE included patients with metastatic disease for whom overall survival data were relatively mature, and those with non-metastatic disease in whom there were relatively few deaths and overall survival benefits remained uncertain. The analyses provided the first formal economic evaluation of the use of docetaxel in this context.

Development of the STM and underlying MSM required consideration of which clinical events or treatment processes to model, and how these should be linked. The process of model conceptualisation was informed by a review of observational data, clinical guidelines and clinical advice. This identified a series of health states which were distinct in terms of mortality risks, HRQoL and (costly) subsequent treatment. The model structure (Paper 5, Figure 1) captures initial treatment failure (development of 'castrate resistant' prostate cancer (CRPC)) at which point individuals enter one of four CRPC states describing the presence and severity of metastases, transitions between these increasingly severe levels of metastatic CRPC; and death due to prostate cancer and other causes. Although a number of trials were relevant to the decision problem the base case analysis focused on evidence from STAMPEDE as this was the largest trial, was UK based, provided detailed data on resource use and HRQoL, and provided IPD. Sensitivity analyses were conducted to explore the impact of including additional evidence.

MSM was used to estimate the rate at which individuals moved through health states. The clinical analysis focused on Failure Free Survival (FFS)[6] and overall survival endpoints. Development of the MSM required a detailed understanding and reshaping of the clinical outcome data from STAMPEDE so that there was a time-to-event for each transition within the model structure. Parametric models were then fitted to the individual transitions. A joint model was fitted across all events considered within the clinical analysis as treatment failure, and covariates were included to allow event rates to differ according to patient baseline characteristics and type of treatment failure. This joint modelling was considered preferred to statistical modelling of individual transitions as it provided sufficient event data to support modelling of a complex baseline hazard; a time-varying treatment effect; and inclusion of multiple baseline characteristics.

Evidence on transitions beyond the point of treatment failure were sparser. Standard parametric functions were therefore used to model these transitions, and they included only two covariates: treatment allocation and time to treatment failure. Time to treatment failure was included as a covariate in subsequent transition models to ensure adequate correspondence over the trial follow-up between the observed overall survival data and the predictions from the MSM. This "history dependence" in the transition from progression/treatment failure to death has been found in previous studies (Majer *et al.*, 2022; Pan *et al.*, 2018; Smare *et al.*, 2019; Williams *et al.*, 2016) and suggests that simpler models which ignore patient history when predicting mortality may not adequately reflect the underlying disease process, and may not, therefore, represent a suitable basis for predicting within-trial and extrapolated overall survival. The decision model uses a patient level simulation approach as this approach can be simpler to implement than cohort simulation when event rates vary over time and according to patient history (Davis *et al.*, 2014).

Explicitly modelling the biological and treatment processes using STM and MSM provided a number of advantages. Long-term outcomes for patients with non-metastatic disease at baseline were simulated using evidence from patients who were metastatic at baseline. Specifically, for non-metastatic patients who developed metastatic CRPC, mortality transition probabilities were generalised from patients who were metastatic at baseline and subsequently developed CRPC. This generalisation of evidence was supported by literature and clinical advice, and made the best use of the available evidence by utilising the rich data available on FFS for non-metastatic patients, and using data from metastatic patients who had received upfront docetaxel or standard of care from the point of treatment failure until death. Had a PSM approach been applied, overall survival

[6] This was defined as a composite endpoint of time to biochemical failure, local progression, local/lymph node/metastatic progression or death from prostate cancer.

extrapolations amongst non-metastatic patients would have relied on immature overall survival evidence introducing additional uncertainty into the decision-making process.

The STM also allowed the long-term cost implications of docetaxel initiation to be appropriately characterised. An important cost driver is whether patients develop CRPC as this typically determines the use of expensive subsequent therapies (Paper 5, Table 1). By explicitly modelling whether a patient's first event is treatment failure or death the model can account for the proportion of patients who are expected to experience CRPC and require further lines of treatment. This is not feasible within PSM which does not quantify which of those patients making an FFS transition experience non-fatal failure (as oppose to death).

The approach taken allowed heterogeneity in FFS and overall survival to be reflected. Heterogeneity in the baseline risk of FFS was modelled by including patient characteristics within the FFS parametric model (Paper 5, Supplementary Tables 5-6). Heterogeneity in overall survival reflects heterogeneity in FFS (as overall survival is the sum of time spent in all health states), and the modelled relationships between FFS and subsequent transitions. An analysis of heterogeneity using PSM would have required direct estimation of the impact of patient characteristics on overall survival. Though this would have been feasible for metastatic patients it is likely to have lacked robustness for non-metastatic patients due to the low number of deaths (James *et al.*, 2016).

The STM structure ensured that sensitivity analyses exploring uncertainties in the disease and treatment process fully reflected effects on overall survival. For example, during the trial an important change in practice was observed with many patients with non-metastatic disease now expected to receive radiotherapy alongside hormone therapy initiation. This was simulated within the model by including planned radiotherapy as a covariate in the FFS analysis and re-running the model assuming all patients with non-metastatic disease had planned radiotherapy. The effects on overall survival were propagated through the model via the mechanisms described above. Again, this would have been challenging within a PSM due to the paucity of evidence with which to estimate direct effects of radiotherapy on overall survival.

A systematic review and meta-analysis identified 15 other RCTs that were relevant to the decision problem (Vale *et al.*, 2016). It was therefore important to assess the effects on cost-effectiveness of including relative treatment effect estimates based on this broader body of evidence. This was achieved within the model by adjusting the docetaxel FFS and overall survival curves predicted by the MSM and STM to reflect the treatment effect estimates from the meta-analysis. This provided revised estimates of time spent in the failure free state and across the CRPC states. The

proportionate change in time spent across the CRPC states was assumed to apply equally to each individual CRPC state.

This case study confirmed some potential limitations identified within Paper 4 of using STM informed by MSM. Both the decision model and parametric survival modelling were significantly more complex than a PSM. This had important implications for the resources required to develop and validate analyses. The wider range of choices regarding model structure and the parameterisation of the survival models also potentially increase the risk for "gaming" where there are commercial or other vested interests in the results of the evaluation. In addition, although the explicit modelling of the disease process increased transparency around the mechanisms driving the model results, these mechanisms were complex and driven by many parameters. Though the results were reviewed by clinical experts, the extent to which they could fully engage with assessing their validity is debatable.

The findings of this analysis informed NICE recommendations that clinicians offer docetaxel to metastatic patients and discuss docetaxel as an option for non-metastatic patients within the eligible population (National Institute for Health and Care Excellence, 2019).

Conclusions

Paper 4 identifies the assumptions underpinning PSM, highlights the importance of assessing the credibility of PSM state membership predictions in the context of all available endpoint data, and identifies STM as an important alternative method that requires further development. Paper 5 implements STM informed by MSM as the primary vehicle for an evaluation. This highlighted its advantages in enabling generalisation of evidence to support extrapolation, more accurate assessment of costs, and allowing both uncertainty and heterogeneity to be reflected in overall survival despite the immaturity of the overall survival evidence. These advantages must be weighed against the significant additional complexity of the approach.

5. Discussion

This thesis makes a number of novel and important contributions to advancing the methods, application and interpretation of cost-effectiveness analysis informed by survival data. The coherence and methodological importance of these developments for improving health care decision making are illustrated in the context of three case studies where decision models were developed to assess cost-effectiveness informed by a range of survival analytic methods. As highlighted within the body of the thesis, the work presented has directly informed NICE Technology Appraisal recommendations; UK and international clinical guidelines and international methods guidance. This further demonstrates the value of these developments for analysts and decision makers, and for improving overall population health.

The approaches developed allowed heterogeneity across patients in clinical event risk and expected benefits of treatment to be reflected within cost-effectiveness analyses in a range of contexts including when data is available from a single trial or multiple trials, and when the model and survival analysis are structured around a single, or multiple, interrelated events. This allowed cost-effectiveness results to reflect important clinical differences between patients, and, optimised recommendations to be made using methodologically robust and transparent analyses.

The assessment and implementation of different decision modelling approaches within this thesis highlights the importance of more explicitly linking the model structure to the required underlying survival analyses. Decision model structures that directly model overall survival (the PSM approach) were used in two case studies. This showed that as PSM disconnects overall survival from other modelled disease and treatment processes, this can limit the ability of evidence on intermediate outcomes to inform overall survival extrapolations and limit exploration of key extrapolation uncertainties. In contrast, the third case study used a STM approach requiring the development of a more complex survival analysis approach (MSM). This allowed a more explicit model of the disease and treatment process to be developed allowing fuller use of the overall evidence base and exploration of heterogeneity and extrapolation uncertainties. This work highlights how these approaches differ in data requirements, assumptions, ability to incorporate external data, and types of uncertainty they can most readily characterise. Rather than the PSM being the default approach, this work suggests that model choice should be based on a careful process of model conceptualisation that considers the nature of the direct and external evidence available, and the nature of the key uncertainties within the decision problem at hand. Further research could usefully explore the utility of STM and MSM for decision making in the context of other case studies, and explore the implications of different modelling approaches for assessments of the value of further research.

Approaches to the analysis of survival data to inform cost-effectiveness models, and to evidence synthesis of survival data, have evolved significantly since the publication of papers 1-3. Within papers 1 and 3, the external validity of overall survival predictions was assessed by comparing predictions from the survival models with general population mortality. Recent methodological advances and guidance support more explicit incorporation of general population mortality within parametric survival models, for example using excess hazard methods (Rutherford *et al.*, 2020; Sweeting *et al.*, 2023; van Oostrum *et al.*, 2021). Current best practice would also involve consideration of a broader range of external evidence such as additional clinical trial evidence, observational studies or expert clinical opinion, and where feasible the application of statistical methods that explicitly incorporate this evidence within parametric survival modelling (Bullement *et al.*, 2023; Chaudhary *et al.*, 2023; Rutherford *et al.*, 2020; Cope *et al.*, 2019). This additional external evidence is likely to differ from the primary clinical evidence in several ways including relevance to the target population, follow-up duration, and the nature of summary statistics reported. In this context, the flexibility of Bayesian multi-parameter evidence synthesis methods may be especially useful (Chaudhary *et al.*, 2023; Guyot *et al.*, 2017; Che, Green and Baio, 2023). Explicit integration of general population mortality, and identification and integration of other relevant external evidence within the overall survival models in papers 1 and 3, would have improved the validity of the long-term survival predictions.

Since the publication of papers 2 and 3, approaches for conducting NMA of survival IPD have been developed in several ways. Due to computational constraints the IPD NMA presented in these papers was restricted to a fixed effects analysis. Methods are now available that allow for modelling of random treatment effects in this context, in a way that is computationally feasible (Freeman and Carpenter, 2017). Application of these methods would have allowed for reflection and quantification of residual across-trial heterogeneity. This may have improved the fit and credibility of the models, and informed selection of effect modifiers. The inclusion of multiple effect modifiers within the NMA model runs the risk of overfitting and increased parameter uncertainty, though excluding effects runs the risk of missing important characteristics that modify treatment effects (Seo *et al.*, 2021). Alternative methods that "shrink" coefficients towards zero have been shown to perform better in simulation studies (i.e. give better estimates of the true subgroup-specific treatment effects) (Seo *et al.*, 2021). Application of these methods within Paper 2 would have reduced the risk of overfitting and would be expected to reduce uncertainty in the effects of the devices on all-cause mortality. How this would modify the mean treatment effects within each subgroup presented in Paper 2, or the cost-effectiveness results within Paper 3, is difficult to predict.

The novel approaches presented within this thesis often use statistical and decision modelling approaches that are complex, introducing additional burden to those tasked with developing and

validating cost-effectiveness models, and interpreting their results. It is therefore important to assess the conditions under which these more complex approaches may or may not be appropriate, in order to ensure the modelling undertaken is proportionate.

Detailed modelling of heterogeneity in baseline risk is likely to be important where average cost-effectiveness within a broader population is close to the cost-effectiveness threshold and/or there is substantial variation in risks of events that are important determinants of morbidity, mortality or resource use. Detailed modelling of heterogeneity in relative treatment effects is likely to be appropriate where effect modification is well supported (i.e. is biologically plausible, supported by a pre-specified rationale and consistent across studies). Under these circumstances IPD meta-analysis or NMA is the gold standard (Riley *et al.*, 2023), though may not always be feasible due to resource constraints or data availability. Methods are available to support an assessment of whether IPD-meta-analysis or NMA is likely to be worthwhile and alternative methods are available (Tudur Smith *et al.*, 2016).

External data is likely to be particularly important where trial evidence is immature, and extrapolation of within-trial trends in event risk is likely to result in high levels of uncertainty. External evidence may take a wide variety of forms and recent studies have shown how external evidence obtained from disease registries; general population mortality; meta-analyses of the relationship between surrogate and final endpoints; indirectly related clinical trials and formal expert elicitation can be integrated with direct trial evidence (Ayers *et al.*, 2022; Batteson *et al.*, 2020; Chaudhary *et al.*, 2023; Cope *et al.*, 2019; Creemers *et al.*, 2023; Creemers *et al.*, 2021). The form of external evidence available may influence the choice of model structure and accompanying survival analysis. For example, in an evaluation of nivolumab for melanoma (Batteson *et al.*, 2020) PSM was applied when using external evidence on the surrogate relationship between recurrence free and overall survival to inform overall survival predictions, whereas STM was used to incorporate external trial evidence on post-recurrence survival. Within the third prostate cancer case study in this thesis, a STM was used to generalise post-failure survival outcomes from patients with metastatic disease at baseline to those with non-metastatic disease at baseline.

Research to date has focused on using external data to inform estimates of baseline survival risk. Economic evaluation results are also typically sensitive to (long-term) treatment effects on overall survival (estimated directly as in PSM or indirectly via multiple parameters as in STMs). Appropriate sources of external evidence relating to the effect of treatment may include data for the same product in a related clinical context; data for a related product in the same clinical context; or expert opinion (Jankovic *et al.*, 2022; Palmer *et al.*, 2023). Further methods research could usefully focus on how external evidence on the effects of treatment over time can be integrated within decision models.

Given the likely limitations of external evidence on the effects of treatment, novel approaches may be warranted. For example, recent mechanistic models describing the dynamics and interrelationship between tumour growth and the immune system, and the effects of immunotherapies on these processes, have shown promise as a vehicle for predicting survival outcomes (Creemers *et al.*, 2023; Creemers *et al.*, 2021). Further work is required to validate these approaches and ascertain their relevance in the context of economic evaluation and reimbursement decision making.

Within Papers 1 and 5, decision analytic modelling approaches were used to reflect the costs and outcomes associated with post-progression therapies. These model elements were parameterised by both aggregate published evidence and bespoke analysis of IPD. A related question that has been addressed in the literature, is how to adjust overall survival evidence from clinical studies where the use of subsequent therapies does not reflect clinical practice (Latimer *et al.*, 2014; Skaltsa *et al.*, 2017). Further research could usefully explore when these alternative approaches are likely to be most appropriate.

Conclusion

This thesis highlights the importance of approaches that more explicitly link decision-analytic model structures and accompanying survival analyses to the underlying disease and treatment processes of interest. This allows for more comprehensive use of evidence, and explicit assessment of the effects of heterogeneity and uncertainty. The work presented highlights how this can be achieved in a wide range of contexts. The choice of when to implement the approaches developed within this thesis will rest on a model conceptualisation process that considers the anticipated importance of heterogeneity in survival outcomes, the nature of direct and external survival evidence, and which aspects of uncertainty it will be important for the modelling approach to reflect.

6. References

Association of British Healthcare Industries (2012) *Implantable cardioverter defibrillators for the treatment of arrhythmias and cardiac resynchronisation therapy for the treatment of heart failure (review of TA95 and TA120)*. Available at: https://www.nice.org.uk/guidance/ta314/resources/arrythmias-icds-heart-failure-cardiac-resynchronisation-evaluation-report4.

Ayers, D., Cope, S., Towle, K., Mojebi, A., Marshall, T. and Dhanda, D. (2022) 'Structured expert elicitation to inform long-term survival extrapolations using alternative parametric distributions: a case study of CAR T therapy for relapsed/ refractory multiple myeloma', *BMC Med Res Methodol,* 22(1), pp. 272.

Batteson, R., Hart, R., Hemstock, M., Gooden, K., Kotapati, S., Roze, S., Lee, D. and Amadi, A. (2020) 'Modelling Survival of Patients Treated with Adjuvant Nivolumab Who Have Melanoma with Lymph Node Involvement or Metastatic Disease After Complete Resection', *Pharmacoecon Open,* 4(2), pp. 343-351.

Bell Gorrod, H., Kearns, B., Stevens, J., Thokala, P., Labeit, A., Latimer, N., Tyas, D. and Sowdani, A. (2019) 'A Review of Survival Analysis Methods Used in NICE Technology Appraisals of Cancer Treatments: Consistency, Limitations, and Areas for Improvement', *Medical Decision Making,* 39(8), pp. 899-909.

Bongers, M., de Ruysscher, D., Oberije, C., Lambin, P., Uyl–de Groot, C. and Coupe´, V. (2016) 'Multistate Statistical Modeling: A Tool to Build a Lung Cancer Microsimulation Model That Includes Parameter Uncertainty and Patient Heterogeneity', *Medical Decision Making,* 36, pp. 86-100.

Briggs, A., Mihaylova, B., Sculpher, M., Hall, A., Wolstenholme, J., Simoons, M., Deckers, J., Ferrari, R., Remme, W. J., Bertrand, M., Fox, K. and Investigators, E. T. (2007) 'Cost effectiveness of perindopril in reducing cardiovascular events in patients with stable coronary artery disease using data from the EUROPA study', *Heart,* 93(9), pp. 1081-6.

Briggs, A., Sculpher, M. and Claxton, K. (2006) 'Decision Modelling for Health Economic Evaluation (1st Edition). Oxford, UK.', *Oxford University Press*, pp. 1-76.

Bullement, A., Cranmer, H. L. and Shields, G. E. (2019) 'A Review of Recent Decision-Analytic Models Used to Evaluate the Economic Value of Cancer Treatments', *Appl Health Econ Health Policy,* 17(6), pp. 771-780.

Bullement, A., Stevenson, M. D., Baio, G., Shields, G. E. and Latimer, N. R. (2023) 'A Systematic Review of Methods to Incorporate External Evidence into Trial-Based Survival Extrapolations for Health Technology Assessment', *Med Decis Making,* 43(5), pp. 610-620.

Castelli, C., Combescure, C., Foucher, Y. and Daures, J. P. (2007) 'Cost-effectiveness analysis in colorectal cancer using a semi-Markov model', *Stat Med,* 26(30), pp. 5557-71.

Chaudhary, M. A., Edmondson-Jones, M., Baio, G., Mackay, E., Penrod, J. R., Sharpe, D. J., Yates, G., Rafiq, S., Johannesen, K., Siddiqui, M. K., Vanderpuye-Orgle, J. and Briggs, A. (2023) 'Use of Advanced Flexible Modeling Approaches for Survival Extrapolation from Early Follow-up Data in two Nivolumab Trials in Advanced NSCLC with Extended Follow-up', *Med Decis Making,* 43(1), pp. 91-109.

Che, Z., Green, N. and Baio, G. (2023) 'Blended Survival Curves: A New Approach to Extrapolation for Time-to-Event Outcomes from Clinical Trials in Health Technology Assessment', *Med Decis Making,* 43(3), pp. 299-310.

Claxton, K. (1999) 'The irrelevance of inference: a decision-making approach to the stochastic evaluation of health care technologies', *J Health Econ,* 18(3), pp. 341-64.

Claxton, K., Palmer, S., Longworth, L., Bojke, L., Griffin, S., McKenna, C., Soares, M., Spackman, E. and Youn, J. (2012) 'Informing a decision framework for when NICE should recommend the use of health technologies only in the context of an appropriately designed programme of evidence development', *Health Technol Assess,* 16(46), pp. 1-323.

Collett, D. (2003) 'Modelling Survival Data in Medical Research', *Chapman and Hall*.

Cope, S., Ayers, D., Zhang, J., Batt, K. and Jansen, J. P. (2019) 'Integrating expert opinion with clinical trial data to extrapolate long-term survival: a case study of CAR-T therapy for children and young adults with relapsed or refractory acute lymphoblastic leukemia', *BMC Med Res Methodol,* 19(1), pp. 182.

Coyle, D., Buxton, M. J. and O'Brien, B. J. (2003) 'Stratified cost-effectiveness analysis: a framework for establishing efficient limited use criteria', *Health Econ,* 12(5), pp. 421-7.

Creemers, J. H. A., Ankan, A., Roes, K. C. B., Schroder, G., Mehra, N., Figdor, C. G., de Vries, I. J. M. and Textor, J. (2023) 'In silico cancer immunotherapy trials uncover the consequences of therapy-specific response patterns for clinical trial design and outcome', *Nat Commun,* 14(1), pp. 2348.

Creemers, J. H. A., Lesterhuis, W. J., Mehra, N., Gerritsen, W. R., Figdor, C. G., de Vries, I. J. M. and Textor, J. (2021) 'A tipping point in cancer-immune dynamics leads to divergent immunotherapy responses and hampers biomarker discovery', *J Immunother Cancer,* 9(5).

Davis, S., Martyn-St James, M., Sanderson, J., Stevens, J., Goka, E., Rawdin, A., Sadler, S., Wong, R., Campbell, F., Stevenson, M., Strong, M., Selby, P. and Gittoes, N. (2016) 'A systematic review and economic evaluation of bisphosphonates for the prevention of fragility fractures', *Health Technol Assess,* 20(78), pp. 1-406.

Davis, S., Stevenson, M., Tappenden, P. and Wailoo, A. (2014) 'NICE DSU Technical Support Document 15: Cost-Effectiveness Modelling Using Patient-Level Simulation', *NICE Decision Support Unit*.

Debray, T. P., Schuit, E., Efthimiou, O., Reitsma, J. B., Ioannidis, J. P., Salanti, G. and Moons, K. G. (2018) 'An overview of methods for network meta-analysis using individual participant data: when do benefits arise?', *Stat Methods Med Res,* 27(5), pp. 1351-1364.

Dias, S., Ades, A. E., Welton, N. J., Jansen, J. P. and Sutton, A. J. (2018) *Network meta-analysis for decision-making.* John Wiley & Sons.

Dias, S., Sutton, A. J., Welton, N. J. and Ades, A. E. (2013) 'Evidence synthesis for decision making 3: heterogeneity--subgroups, meta-regression, bias, and bias-adjustment', *Med Decis Making,* 33(5), pp. 618-40.

Drummond, M. F., Sculpher, M. J., Claxton, K., Stoddart, G. L. and Torrance, G. W. (2015) *Methods for the economic evaluation of health care programmes.* Oxford university press.

Ezekowitz, J. A., O'Meara, E., McDonald, M. A., Abrams, H., Chan, M., Ducharme, A., Giannetti, N., Grzeslo, A., Hamilton, P. G., Heckman, G. A., Howlett, J. G., Koshman, S. L., Lepage, S., McKelvie, R. S., Moe, G. W., Rajda, M., Swiggum, E., Virani, S. A., Zieroth, S., Al-Hesayen, A., Cohen-Solal, A., D'Astous, M., De, S., Estrella-Holder, E., Fremes, S., Green, L., Haddad, H., Harkness, K., Hernandez, A. F., Kouz, S., LeBlanc, M. H., Masoudi, F. A., Ross, H. J., Roussin, A. and Sussex, B. (2017) '2017 Comprehensive Update of the Canadian Cardiovascular Society Guidelines for the Management of Heart Failure', *Can J Cardiol,* 33(11), pp. 1342-1433.

Faria, R., Woods, B., Griffin, S., Palmer, S., Sculpher, M. and Ryder, S. D. (2016) 'Prevention of progression to cirrhosis in hepatitis C with fibrosis: effectiveness and cost effectiveness of sequential therapy with new direct-acting anti-virals', *Aliment Pharmacol Ther,* 44(8), pp. 866-76.

Fox, M., Mealing, S., Anderson, R., Dean, J., Stein, K., Price, A. and Taylor, R. (2007) 'The clinical effectiveness and cost-effectiveness of cardiac resynchronisation (biventricular pacing) for heart failure: systematic review and economic model', *Health Technology Assessment (Winchester, England),* 11(47), pp. iii-iv, ix.

Freeman, S. C. and Carpenter, J. R. (2017) 'Bayesian one-step IPD network meta-analysis of time-to-event data using Royston-Parmar models', *Res Synth Methods,* 8(4), pp. 451-464.

Freeman, S. C., Cooper, N. J., Sutton, A. J., Crowther, M. J., Carpenter, J. R. and Hawkins, N. (2022) 'Challenges of modelling approaches for network meta-analysis of time-to-event outcomes in the presence of non-proportional hazards to aid decision making: Application to a melanoma network', *Stat Methods Med Res,* 31(5), pp. 839-861.

Guyot, P., Ades, A. E., Beasley, M., Lueza, B., Pignon, J. P. and Welton, N. J. (2017) 'Extrapolation of Survival Curves from Cancer Trials Using External Information', *Med Decis Making,* 37(4), pp. 353-366.

Hadwiger, M., Frielitz, F. S., Eisemann, N., Elsner, C., Dagres, N., Hindricks, G. and Katalinic, A. (2021) 'Cardiac Resynchronisation Therapy in Patients with Moderate to Severe Heart Failure in Germany: A Cost-Utility Analysis of the Additional Defibrillator', *Appl Health Econ Health Policy,* 19(1), pp. 57-68.

Haute Autorité de Santé (2020) 'Choices in methods for economic evaluation—HAS', *Methodological Guidance2020*.

Health Information and Quality Authority (2020) *Guidelines for the Economic Evaluation of Health Technologies in Ireland*, Dublin: HIQA.

Hoyle, M., Crathorne, L., Jones-Hughes, T. and Stein, K. (2010) *Bendamustine for the first-line treatment of chronic lymphocytic leukaemia (Binet stage B or C) in patients for whom fludarabine combination chemotherapy is not appropriate: a critique of the submission from Napp*.

Jackson, C., Stevens, J., Ren, S., Latimer, N., Bojke, L., Manca, A. and Sharples, L. (2017) 'Extrapolating Survival from Randomized Trials Using External Data: A Review of Methods', *Med Decis Making,* 37(4), pp. 377-390.

James, N. D., Sydes, M. R., Clarke, N. W., Mason, M. D., Dearnaley, D. P., Spears, M. R., Ritchie, A. W., Parker, C. C., Russell, J. M., Attard, G., de Bono, J., Cross, W., Jones, R. J., Thalmann, G., Amos, C., Matheson, D., Millman, R., Alzouebi, M., Beesley, S., Birtle, A. J., Brock, S., Cathomas, R., Chakraborti, P., Chowdhury, S., Cook, A., Elliott, T., Gale, J., Gibbs, S., Graham, J. D., Hetherington, J., Hughes, R., Laing, R., McKinna, F., McLaren, D. B., O'Sullivan, J. M., Parikh, O., Peedell, C., Protheroe, A., Robinson, A. J., Srihari, N., Srinivasan, R., Staffurth, J., Sundar, S., Tolan, S., Tsang, D., Wagstaff, J., Parmar, M. K. and investigators, S. (2016) 'Addition of docetaxel, zoledronic acid, or both to first-line long-term hormone therapy in prostate cancer (STAMPEDE): survival results from an adaptive, multiarm, multistage, platform randomised controlled trial', *Lancet,* 387(10024), pp. 1163-77.

Jankovic, D., Payne, K., Kanaan, M. and Bojke, L. (2022) 'Eliciting uncertainty for complex parameters in model-based economic evaluations: quantifying a temporal change in the treatment effect', *Int J Technol Assess Health Care,* 38(1), pp. e21.

Kaltenthaler, E., Tappenden, P., Paisley, S. and Squires, H. (2011) *NICE DSU Technical Support Document 13: Identifying and Reviewing Evidence to Inform the Conceptualisation and Population of Cost-Effectiveness Models*, London: National Institute for Health and Care Excellence (NICE).

Latimer, N. (2011) 'NICE DSU Technical Support Document 14: Survival Analysis For Economic Evaluations Alongside Clinical Trials - Extrapolation With Patient-Level Data ', *NICE Decision Support Unit*.

Latimer, N. R., Abrams, K. R., Lambert, P. C., Crowther, M. J., Wailoo, A. J., Morden, J. P., Akehurst, R. L. and Campbell, M. J. (2014) 'Adjusting survival time estimates to account for treatment switching in randomized controlled trials--an economic evaluation context: methods, limitations, and recommendations', *Med Decis Making,* 34(3), pp. 387-402.

Latimer, N. R. and Adler, A. I. (2022) 'Extrapolation beyond the end of trials to estimate long term survival and cost effectiveness', *BMJ Medicine,* 1(1).

Majer, I., Kroep, S., Maroun, R., Williams, C., Klijn, S. and Palmer, S. (2022) 'Estimating and Extrapolating Survival Using a State-Transition Modeling Approach: A Practical Application in Multiple Myeloma', *Value Health,* 25(4), pp. 595-604.

McDonagh, T. A., Metra, M., Adamo, M., Gardner, R. S., Baumbach, A., Böhm, M., Burri, H., Butler, J., Čelutkienė, J., Chioncel, O., Cleland, J. G. F., Coats, A. J. S., Crespo-Leiro, M. G., Farmakis, D., Gilard, M., Heymans, S., Hoes, A. W., Jaarsma, T., Jankowska, E. A., Lainscak, M., Lam, C. S. P., Lyon, A. R., McMurray, J. J. V., Mebazaa, A., Mindham, R., Muneretto, C., Francesco Piepoli, M., Price, S., Rosano, G. M. C., Ruschitzka, F. and Kathrine Skibelund, A. (2021) '2021 ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure', *Eur Heart J,* 42(36), pp. 3599-3726.

National Institute for Health and Care Excellence (2011) *Bendamustine for the first-line treatment of chronic lymphocytic leukaemia: Technology appraisal guidance*. Available at: www.nice.org.uk/guidance/ta216.

National Institute for Health and Care Excellence (2013) 'NICE Process and Methods Guides', *Guide to the Methods of Technology Appraisal 2013*. London: National Institute for Health and Care Excellence (NICE).

National Institute for Health and Care Excellence (2014) 'Implantable cardioverter defibrillators and cardiac resynchronisation therapy for arrhythmias and heart failure', *NICE Technology Appraisal Guidance*, pp. 314.

National Institute for Health and Care Excellence (2019) *Prostate cancer: diagnosis and management*. Available at: www.nice.org.uk/guidance/ng131.

National Institute for Health and Care Excellence (2022) *NICE health technology evaluations: the manual*. Available at: https://www.nice.org.uk/process/pmg36/chapter/introduction-to-health-technology-evaluation.

Neyt, M., García-Pérez, L., Johansson, P., Midy, F. and Teljeur, C. 2020. Practical considerations when critically assessing economic evaluations. Guidance document. European Network for Health Technology Assessment (EUnetHTA).

Nikolaidis, G. F., Woods, B., Palmer, S. and Soares, M. O. (2021) 'Classifying information-sharing methods', *BMC Med Res Methodol,* 21(1), pp. 107.

Palmer, S., Borget, I., Friede, T., Husereau, D., Karnon, J., Kearns, B., Medin, E., Peterse, E. F. P., Klijn, S. L., Verburg-Baltussen, E. J. M., Fenwick, E. and Borrill, J. (2023) 'A Guide to Selecting Flexible Survival Models to Inform Economic Evaluations of Cancer Immunotherapies', *Value Health,* 26(2), pp. 185-192.

Pan, F., Reifsnider, O., Zheng, Y., Proskorovsky, I., Li, T., He, J. and Sorensen, S. V. J. V. i. H. (2018) 'Modeling clinical outcomes in prostate cancer: application and validation of the discrete event simulation approach', 21(4), pp. 416-422.

Ponikowski, P., Voors, A. A., Anker, S. D., Bueno, H., Cleland, J. G. F., Coats, A. J. S., Falk, V., González-Juanatey, J. R., Harjola, V. P., Jankowska, E. A., Jessup, M., Linde, C., Nihoyannopoulos, P., Parissis, J. T., Pieske, B., Riley, J. P., Rosano, G. M. C., Ruilope, L. M., Ruschitzka, F., Rutten, F. H. and van der Meer, P. (2016) '2016 ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure: The Task Force for the diagnosis and treatment of acute and chronic heart failure of the European Society of Cardiology (ESC)Developed with the special contribution of the Heart Failure Association (HFA) of the ESC', *Eur Heart J,* 37(27), pp. 2129-2200.

Riley, R. D., Dias, S., Donegan, S., Tierney, J. F., Stewart, L. A., Efthimiou, O. and Phillippo, D. M. (2023) 'Using individual participant data to improve network meta-analysis projects', *BMJ Evid Based Med,* 28(3), pp. 197-203.

Rutherford, M., Lambert, P., Sweeting, M., Pennington, R., Crowther, M. J., Abrams, K. R. and Latimer, N. R. (2020) 'NICE DSU Technical Support Document 21: Flexible methods for survival analysis', *Department of Health Sciences, University of Leicester, Leicester, UK*, pp. 1-97.

Scottish Medicines Consortium (2011) *bendamustine hydrochloride 25mg, 100mg powder for solution for infusion (Levact®) SMC No. (694/11)*. Available at: https://www.scottishmedicines.org.uk/medicines-advice/bendamustine-hydrochloride-levact-fullsubmission-69411/.

Seo, M., White, I. R., Furukawa, T. A., Imai, H., Valgimigli, M., Egger, M., Zwahlen, M. and Efthimiou, O. (2021) 'Comparing methods for estimating patient-specific treatment effects in individual patient data meta-analysis', *Stat Med,* 40(6), pp. 1553-1573.

Shah, D., Lu, X., Paly, V. F., Tsintzos, S. I. and May, D. M. (2020) 'Cost-effectiveness analysis of implantable cardiac devices in patients with systolic heart failure: a US perspective using real world data', *J Med Econ,* 23(7), pp. 690-697.

Siebert, U., Alagoz, O., Bayoumi, A., Jahn, B., Owens, D., Cohen, D. and Kuntz, K. (2012) 'State-transition modeling: a report of the ISPOR-SMDM modeling good research practices task force-3', *Value in Health,* 15(6), pp. 812-820.

Skaltsa, K., Ivanescu, C., Naidoo, S., Phung, D., Holmstrom, S. and Latimer, N. R. (2017) 'Adjusting Overall Survival Estimates after Treatment Switching: a Case Study in Metastatic Castration-Resistant Prostate Cancer', *Target Oncol,* 12(1), pp. 111-121.

Smare, C., Lakhdari, K., Doan, J., Posnett, J. and Johal, S. (2019) 'Evaluating Partitioned Survival and Markov Decision-Analytic Modeling Approaches for Use in Cost-Effectiveness Analysis: Estimating and Comparing Survival Outcomes', *Pharmacoeconomics*.

Sweeting, M. J., Rutherford, M. J., Jackson, D., Lee, S., Latimer, N. R., Hettle, R. and Lambert, P. C. (2023) 'Survival Extrapolation Incorporating General Population Mortality Using Excess Hazard and Cure Models: A Tutorial', *Med Decis Making,* 43(6), pp. 737-748.

Tai, T. A., Latimer, N. R., Benedict, A., Kiss, Z. and Nikolaou, A. (2021) 'Prevalence of Immature Survival Data for Anti-Cancer Drugs Presented to the National Institute for Health and Care Excellence and Impact on Decision Making', *Value Health,* 24(4), pp. 505-512.

Tudur Smith, C., Marcucci, M., Nolan, S. J., Iorio, A., Sudell, M., Riley, R., Rovers, M. M. and Williamson, P. R. (2016) 'Individual participant data meta-analyses compared with meta-analyses based on aggregate data', *Cochrane Database Syst Rev,* 9(9), pp. MR000007.

Vale, C. L., Burdett, S., Rydzewska, L. H. M., Albiges, L., Clarke, N. W., Fisher, D., Fizazi, K., Gravis, G., James, N. D., Mason, M. D., Parmar, M. K. B., Sweeney, C. J., Sydes, M. R., Tombal, B., Tierney, J. F. and Group, S. T. S. (2016) 'Addition of docetaxel or bisphosphonates to standard of care in men with localised or metastatic, hormone-sensitive prostate cancer: a systematic review and meta-analyses of aggregate data', *Lancet Oncol,* 17(2), pp. 243-256.

van Oostrum, I., Ouwens, M., Remiro-Azocar, A., Baio, G., Postma, M. J., Buskens, E. and Heeg, B. (2021) 'Comparison of Parametric Survival Extrapolation Approaches Incorporating General Population Mortality for Adequate Health Technology Assessment of New Oncology Drugs', *Value Health,* 24(9), pp. 1294-1301.

Ward, T., Medina-Lara, A., Mujica-Mota, R. E. and Spencer, A. E. (2021) 'Accounting for Heterogeneity in Resource Allocation Decisions: Methods and Practice in UK Cancer Technology Appraisals', *Value Health,* 24(7), pp. 995-1008.

Welton, N. J., Sutton, A. J., Cooper, N., Abrams, K. R. and Ades, A. (2012) *Evidence synthesis for decision making in healthcare.* John Wiley & Sons.

Williams, C., Lewsey, J., Mackay, D. and Briggs, A. (2016) '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,* 37(4), pp. 427-439.

Woods, B., Sideris, E., Palmer, S., Latimer, N. and Soares, M. (2017) 'NICE DSU Technical Support Document 19. Partitioned Survival Analysis for Decision Modelling in Health Care: A Critical Review ', *NICE Decision Support Unit*.

7. Appendix 1: published papers

Appendix 1.1: Woods B, Hawkins N, Dunlop W, O'Toole A, Bramham-Jones S. Bendamustine versus chlorambucil for the first-line treatment of chronic lymphocytic leukemia in England and Wales: a cost-utility analysis. Value in health 2012; 15(5):759-70.
DOI: https://doi.org/10.1016/j.jval.2012.03.1389

Appendix 1.2: Woods, B., Hawkins, N., Mealing, S., Sutton, A., Abraham, W.T., Beshai, J.F., Klein, H., Sculpher, M., Plummer, C.J. and Cowie, M.R. Individual patient data network meta-analysis of mortality effects of implantable cardiac devices. Heart 2015; 101(22):1800-6.
DOI: http://dx.doi.org/10.1136/heartjnl-2015-307634

Appendix 1.3: Mealing, S., Woods, B., Hawkins, N., Cowie, M.R., Plummer, C.J., Abraham, W.T., Beshai, J.F., Klein, H., and Sculpher, M. Cost-effectiveness of implantable cardiac devices in patients with systolic heart failure. Heart 2016; 102(21):1742-9.
DOI: http://dx.doi.org/10.1136/heartjnl-2015-308883

Appendix 1.4: Woods, B. S., Eleftherios, S., Palmer, S., Latimer, N., and Soares, M. Partitioned Survival and State Transition Models for Healthcare Decision Making in Oncology: Where Are We Now? Value in Health 2020; 23(12): 1613-21.
DOI: https://doi.org/10.1016/j.jval.2020.08.2094

Appendix 1.5: Woods, B.S., Sideris E., Sydes M.R., Gannon, M.R., Parmar, M.K.B., Alzouebi, M. et al. Addition of Docetaxel to First-line Long-term Hormone Therapy in Prostate Cancer (STAMPEDE): Modelling to Estimate Long-term Survival, Quality-adjusted Survival, and Cost-effectiveness. European Urology Oncology 2018; 1(6):449-58.
DOI: https://doi.org/10.1016/j.euo.2018.06.004