
Article:
Bell Gorrod, H. orcid.org/0000-0001-8054-8073, Kearns, B. orcid.org/0000-0001-7730-668X, Stevens, J. orcid.org/0000-0002-9867-7209 et al. (5 more authors) (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. ISSN 0272-989X

https://doi.org/10.1177/0272989X19881967

TITLE: A review of survival analysis methods used in NICE technology appraisals of cancer treatments: Consistency, limitations and areas for improvement

AUTHORS: Helen Bell Gorrod[1], Ben Kearns[1], John Stevens[1], Praveen Thokala[1], Alexander Labeit[21], Nicholas Latimer[1], David Tyas[3], Ahmed Sowdani[3]

[1] School of Health and Related Research (ScHARR), University of Sheffield

[2] Saw Swee Hock School of Public Health, National University of Singapore

[3] Bristol-Myers Squibb

Word count; 3965

Tables and figures; Table 1: Compliance with the TSD recommendations for extrapolation of OS

Appendix; Table A1: Extraction form section headings

Acknowledgements; We would like to thank Dr Ruth Wong for her assistance with extracting the Technology Appraisal documents from the NICE website.

Declaration of Conflicting Interests: The work on this review was partly funded by Bristol-Myers Squibb and the National Institute for Health Research. NRL and BK are supported by the National Institute for Health Research (NIHR Post-Doctoral Fellowship, Dr Nicholas Latimer, PDF-2015-08-022 and NIHR Doctoral Fellowship, Mr Ben Kearns, DRF-2016-09-119). NRL is supported by Yorkshire Cancer Research (Award reference number S406NL). Ben Kearns was supported in the preparation/submission of this paper by the HEOM Theme of the NIHR CLAHRC Yorkshire and Humber. www.clahrc-yh.nihr.ac.uk. The views expressed in this publication are those of the author(s) and not necessarily those of the NHS, the National Institute for Health Research, or the Department of Health and Social Care. NL authored Technical Support Document 14 for the National Institute for Health and Care Excellence (NICE) Decision Support Unit (DSU), which is referred to in this paper. HBG, BK, JWS, PT, and NL have worked on consultancy projects for a number of pharmaceutical companies. JWS is also a shareholder of AstraZeneca, GlaxoSmithKline and Takeda Pharmaceutical Company Ltd.

Abstract

Objectives: In June 2011, the National Institute for Health and Care Excellence (NICE) Decision Support Unit published a Technical Support Document (TSD) providing recommendations on survival analysis for NICE technology appraisals (TAs). Survival analysis outputs are influential inputs into economic models estimating the cost-effectiveness of new cancer treatments. Hence, it is important that systematic and justifiable model selection approaches are used. This study investigates the extent to which the TSD recommendations have been followed since its publication.

Methods: We reviewed NICE cancer TAs completed between July 2011 and July 2017. Information on survival analyses undertaken and associated critiques for overall survival (OS) and progression-free survival were extracted from the company submissions, Evidence Review Group (ERG) reports and final appraisal determination documents.

Results: Information was extracted from 58 TAs. Only four (7%) followed all TSD recommendations for OS outcomes. The vast majority (91%) compared a range of common parametric models and assessed their fit to the data (86%). Only a minority of TAs included an assessment of the shape of the hazard function (38%) or proportional hazards assumption (40%). Validation of the extrapolated portion of the survival function using external data was attempted in a minority of TAs (40%). Extrapolated survival functions were frequently criticised by ERGs (71%).

Conclusions: Survival analysis within NICE TAs remains sub-optimal, despite publication of the TSD. Model selection is not undertaken in a systematic way resulting in inconsistencies between TAs. More attention needs to be given to assessing hazard functions and validation of extrapolated survival functions. Novel methods not described in the TSD have been used, particularly in the context of immuno-oncology, suggesting that an updated TSD may be of value.

Word count; 275 (max. 275)

Introduction

Economic evaluation requires a comparison of the incremental costs and health effects of competing interventions. Health effects are typically defined in terms of quality adjusted life-years, which involves the analysis of patient time-to-event outcomes such as death and progression-free survival in oncology. In order to avoid delays in obtaining marketing authorisation and reimbursement approval, and because it is usually impractical to run trials with an indefinite length, the timeframe of trials is often shorter than the desired time horizon for economic evaluation. In these cases it is necessary to extrapolate survival functions beyond the duration of a trial. Different assumptions about, and approaches to modelling, the underlying hazard function and resulting survival function can affect the estimates of population mean benefit and incremental cost-effectiveness ratios (ICERs),[1, 2, 3] particularly when there is a high proportion of unobserved events.[3] Consequently, demonstrating that appropriate survival models have been selected is important as this increases the likelihood that good funding decisions are made.

Extrapolation of hazard and survival functions is particularly relevant in the context of cancer health technology assessments (HTAs) because a proportion of overall survival (OS) events are often censored (i.e. have not yet been observed) at the end of trial follow-up. In the context of cancer HTAs, we require an estimate of the population mean survival over a time horizon that captures all differences in health benefit, which is usually lifetime. [4, 5, 6, 7, 8] In order to adopt a lifetime perspective, assumptions must be made about the hazard and survival functions beyond the collection of sample data.

There are many approaches that can be taken to extrapolate hazard and survival functions. Commonly, parametric models are fitted to the sample data to provide a lifetime extrapolation. However, model selection should not depend only on the goodness-of-fit of the model to the sample data but also on the clinical plausibility of the extrapolated hazard and survival functions. Different models can provide very different extrapolated hazard and survival functions, and it is essential that the model is selected carefully to ensure that the cost effectiveness analysis (CEA) is based upon a plausible estimate of population mean survival.

Within England the National Institute for Health and Care Excellence (NICE) conducts technology appraisals (TAs) and provides guidance to the National Health Service. A typical TA involves a company submission providing evidence on the clinical and cost-effectiveness of the technology appraised. This evidence is assessed by an independent Evidence Review Group (ERG), which produces its own report. These reports are considered by the NICE Appraisal Committee, which ultimately produces a final appraisal determination (FAD). The FAD document highlights concerns about the analysis or methodological shortcomings of the submission and reports the outcome of the appraisal. There are three possible outcomes of the FAD: the technology is or is not recommended as a treatment option or (for cancer appraisals only) it is recommended for use through the Cancer Drugs Fund. The Cancer Drugs Fund is a managed access fund. It is an option when both the clinical and cost-effectiveness evidence are too uncertain for a positive recommendation, but the technology has the potential to be cost-effective with additional data collection. Further details on the process of NICE TAs is available on their website.[9]

The NICE Decision Support Unit (DSU) published a Technical Support Document (TSD 14) in June 2011 to address problems with the inconsistency and application of survival analysis methods

incorporated in TAs.[10] The TSD provided recommendations for choosing the appropriate survival modelling approach based on an assessment of the internal and external validity of the extrapolated hazard and survival functions. These recommendations intended to improve the quality and consistency of survival modelling in NICE submissions. However, it is not clear to what extent these recommendations are being followed in submissions to NICE since the TSD was published.

This paper provides a review of models used to extrapolate hazard and survival functions in NICE Cancer TAs which commenced following the publication of the TSD. It aims to determine whether the TSD has been followed and has led to an improvement in terms of consistent and systematic application of assessments to inform the selection of appropriate survival extrapolation in NICE Cancer TAs, and highlight areas where there is need for further progress.

Methods

The purpose of this review was to identify the number of Single Technology Appraisals (STAs) that followed recommendations TSD 14. TSD 14 made recommendations regarding the assessment of internal and external validity of survival models, and incorporation of uncertainty into CEA.

The methods used for the review followed similar reviews of NICE TAs.[10, 11] The scope was restricted to completed NICE STAs for cancer treatments which commenced between 1st July 2011 and 30th June 2017. STAs that involved company submissions made prior to 1st July 2011 were excluded from our review because it would not have been possible for these to follow the TSD recommendations. Multiple Technology appraisals (MTAs) were considered to be out of scope because those submissions rely on network meta-analysis, which is not directly relevant to TSD 14. The STAs that were within scope were identified from the NICE website [9]. Company submissions, ERG reports, and FAD documents were available for each STA and were obtained by information specialists and stored on a Mendeley online database [12]. When a STA had been updated by a more recent appraisal, evidence from both sources was extracted and counted as one submission, to avoid double counting.

A data extraction form was created by the review team to ensure that the necessary information was extracted to meet the aims and objectives of the review. Reviewers collected information on the compliance of company submissions and assessment group reports with the recommendations of the TSD, in terms of the extrapolation of OS and, where available, progression-free survival (PFS). Section headings of the data extraction form corresponded to the key recommendations from TSD 14[10] and align with the headings in Table 1. Information was also extracted on the extrapolation approaches used. In addition, information was extracted from ERG reports regarding criticisms of the models used to extrapolate hazard and survival functions applied in company submissions and regarding alternative approaches suggested or undertaken. From the FAD, issues relating to the extrapolation of survival functions that affected the final decision were highlighted. Most of the extraction required a binary 'yes/no' response, although free text was extracted where additional detail was thought to be useful. Further details of the extraction form are provided in the appendix.

The TAs were divided between 4 reviewers, with each reviewer following the same pre-specified process for data extraction as outlined above. Alongside the binary yes/no responses and free text, page numbers of the relevant sections in the TA documents were also noted, to support the data extraction. All reviewers met after completing two reviews each to ensure that they were extracting the information consistently. At the end, 20% of each reviewer's TAs were allocated for second review by two other reviewers (i.e. 10% each) to check the quality of the data extraction. Final checks of each review were made by the lead reviewer to confirm the accuracy of the data extractions.

Results

Compliance with TSD recommendations

There were 58 STAs within the scope of the review.[13-71] The first company submission was submitted in October 2011, 4 months after the publication of the TSD. All 58 of the STAs undertook an extrapolation of OS, and 47 of the STAs also undertook an extrapolation of PFS. Despite 39 (67%)

of the STAs referencing TSD 14, only 4 (7%) STAs fully followed all of its recommendations for the extrapolation of OS, and none of the TAs fully followed the recommendations for the extrapolation of PFS. Table 1 provides a summary of assessments made in relation to each recommendation.

Table 1: Compliance with TSD 14 recommendations for extrapolation of OS

	Number (%) of TAs					
	PFS			OS		
	N=22	N=25	N=47	N=24	N=34	N=58
	Jul 2011-2014	Jul 2014-2017	Total	Jul 2011-2014	Jul 2014-2017	Total
1) Internal validity	2 (9%)	11 (44%)	13 (28%)	5 (21%)	10 (29%)	15 (26%)
1(a) Assessed shape of the hazard	8 (36%)	13 (52%)	21 (45%)	10 (42%)	12 (35%)	22 (38%)
1(b) Assessed proportional hazards	7 (32%)	13 (52%)	20 (43%)	8 (33%)	15 (44%)	23 (40%)
1(c) Assessed the relative goodness-of-fit using AIC and/or BIC	19 (86%)	23 (92%)	42 (89%)	20 (83%)	30 (88%)	50 (86%)
1(d) Assessed the absolute goodness of fit	16 (73%)	22(88%)	38 (81%)	16 (67%)	29 (85%)	45 (78%)
2) External validity	2 (9%)	2 (8%)	4 (9%)	4 (17%)	7 (21%)	11 (19%)
2(a) Validated extrapolation with external data	3 (14%)	6 (24%)	9 (19%)	8 (33%)	15 (44%)	23 (40%)
2(b) Validated extrapolation according to clinical plausibility	7 (32%)	10 (40%)	17 (36%)	9 (38%)	19 (56%)	28 (48%)
3) Other						
3(a)Considered structural uncertainty in sensitivity analyses	17 (77%)	19 (76%)	36 (77%)	18 (75%)	24 (71%)	42 (72%)

OS: overall survival, PFS: progression-free survival, TA: technology appraisal. The internal validity row (1) indicates the number (%) of TAs that performed assessments 1(a), 1(b), 1(c) and 1(d). The external validity row (2) indicates the number (%) of TAs that validated 2(a) and 2(b)

Internal validity of OS model fit was fully assessed in 15 (26%) TAs and for PFS model fit in 13 (28%) TAs. Full assessment of internal validity included assessing the shape of the hazard function, the plausibility of the proportional hazards assumption, and the goodness-of-fit of the model to the data. Proportional hazards were assessed for OS in 23 (38%) TAs and for PFS in 20 (43%) TAs. Of those that did not assess proportional hazards, 5 (14%) fitted proportional hazard models for OS and 6 (22%) fitted proportional hazard models for PFS. The shape of the hazard function was formally assessed in 22 (38%) TAs for OS and 21 (45%) for PFS. In 50 (86%) of TAs the relative goodness-of-fit of different parametric models for OS was assessed using AIC and/or BIC and in 42 (89%) TAs for PFS. The absolute goodness-of-fit of the OS to the sample data was assessed through visual inspection or using a test of Cox-Snell residuals in 45 (78%) TAs, and in 38 (81%) TAs for PFS. Of these, visual inspection was the more commonly used approach.

With regards to the external validity of extrapolated OS hazard and survival functions, these were validated with both external data and using clinical opinion in the company submission in only 11 (19%) TAs, and in only 4 (9%) TAs in the case of PFS. In 23 (40%) TAs the OS extrapolated hazard and survival functions were validated using external data and in 9 (19%) TAs for PFS. Of the company submissions that did not use external data to validate extrapolated OS hazard and survival functions, there were 2 cases (6%) where the ERG used external data for validation.[50, 72] In 28 (48%) TAs the OS extrapolated hazard and survival functions were validated using clinical opinion and in 17 (36%) TAs for PFS. Additional assessments of the structural uncertainty about the OS extrapolated

survival functions was done using sensitivity analyses in 42 (72%) TAs, and in 36 (77%) TAs in the case of PFS.

To assess if adherence to the TSD had improved over time, the TAs were split into two groups based on time of company submission (1st July 2011 to 30th June 2014 and 1st July 2014 to 30th June 2017). The comparison did not show any overall substantive change in the proportion of submissions that adhered to the TSD recommendations. There was an increase in the proportion of TAs that used external data to validate the model and also the proportion that validated the model based on clinical opinion. More company submissions followed the TSD for the assessment of PFS in the more recent period.

Extrapolation approaches used

A range of different approaches were used to extrapolate survival functions. In 53 (91%) TAs common parametric distributions were used (i.e. exponential, Weibull, gamma, log-normal and log-logistic, which are members of the generalised gamma distribution family, and the Gompertz distribution) to extrapolate OS hazard and survival functions. This was the case with respect to PFS in 43 (91%) TAs. Another widely-used method (n = 14, 23%) was to model the data using piecewise models which split data at change (or cut) points, with extrapolations based on fitting a parametric models to a subset of trial data and/or external data.[15, 20, 40, 54, 68] In 10 submissions the company employed a specific type of piecewise model, known as a hybrid model, which typically split the observed data into two unique time-periods. This uses Kaplan-Meier estimates for the first time-period, with a parametric distribution used for both the second time-period and for generating extrapolations. The hybrid method was also applied in seven immunotherapy (IO) TAs for melanoma,[17, 30, 38, 41, 49, 56, 66] following the precedent for this method set by the first immunotherapy treatment assessed by NICE within this tumour type and reviewed by LRiG.[17]

In three appraisals, where hazard functions were not compatible with commonly used parametric models, other more complex parametric models were used to extrapolate hazard and survival functions. In TA414, a mixture-cure model was used to account for differences in the hazard for different subgroups of patients.[64] Flexible parametric spline-based models were used in TA374 and TA417 to account for changes in the hazard rate over time that cannot be captured using parametric distributions.[66, 72]

ERG critique

ERGs criticised the resulting extrapolated survival functions in 41 (71%) of the 58 STAs. The ERGs identified twenty-three cases where the TSD had been misinterpreted or not followed. These included situations where proportional hazards had been assumed but there had been an incorrect assessment or no assessment of the plausibility of the assumption,[22, 27, 40, 41, 55, 66] and/or where the long-term treatment effect was deemed to be overestimated [23, 25, 26, 29, 36, 39, 43, 54, 56, 57, 61, 63, 65] or the hazard function was not regarded to be clinically plausible.[20, 38, 41, 45, 69] The proportional hazards assumption was mostly assumed when treatment switching adjustment methods had been applied[31, 32, 55, 60], even though it is not a requirement for treatment switching methods. The ERG criticised a submission where the company had applied a HR to a control group survival function to estimate the experimental group survival function, where the base function took the form of an accelerated failure time distribution, including log-normal or

gamma parametric models.[43] ERGs also criticised five cases where external data were not used to validate the extrapolation,[22, 28, 32, 40, 51], three cases where the external data used was deemed to be inappropriate,[38, 62, 68] and five cases where an assessment of goodness-of-fit was given priority over external validation of the extrapolated hazard and survival functions.[30, 31, 39, 40, 63] In one appraisal, the ERG suggested that there should have been more transparency around the elicitation of expert opinion.[33]

ERGs criticised applications of the hybrid method of extrapolation where change-points had been arbitrarily selected rather than based on the shape of the hazard function or when sensitivity to the choice of change-point was not assessed.[15, 19, 54, 69, 68]

The ERGs assessed the sensitivity of ICERs to both alternative parametric models,[15, 26, 51, 52, 53, 66] and more pessimistic assumptions about treatment effects during the extrapolated period (for example, assuming that there is no treatment benefit).[29, 32] ERGs also assessed the impact of alternative change-points as the piecewise models and hybrid models have been criticised for arbitrary choice of change point which can often influence the ICER. [15, 54, 73]

FAD considerations

The FAD documents provided information on the issues of extrapolating hazard and survival functions that were raised during the appraisal committee discussions. Issues relating to the extrapolation of survival functions were discussed in 41 (71%) of the 58 STAs. The main areas of discussion were the suitability of proportional hazards assumptions (n = 8) [13, 27, 33, 35, 38, 41, 55, 60], use of the non-parametric Kaplan-Meier estimator rather than parametric survival functions (n = 6),[15, 26, 40, 46, 47, 54] lack of validation with appropriate external data (n = 2),[28, 45] clinical plausibility of the extrapolated hazard and survival functions (n = 19),[18, 21, 23, 29, 34, 36, 38, 39, 41, 47, 49, 52, 53, 56, 57, 62, 63, 66, 69] structural uncertainty about the most plausible extrapolation (particularly when the duration of follow-up was short relative to patients' lifetime (n = 3) [21, 42, 65]) and treatment switching (n = 5).[16, 19, 20, 22, 55] Situations where there was little or no discussion regarding the extrapolated survival functions occurred when the sensitivity of the ICER had been assessed for a range of plausible models,[24] when the survival function was mature [15, 26] or in cases when the treatment was dominated.[20, 43]

Discussion

Survival models make assumptions about the shape of the underlying hazard function. Therefore, it is important to assess the suitability of the model based on what was observed in the trial period and clinical knowledge of both the disease process and mechanism of action of the treatment.[74] These assessments will also inform what we expect to observe in the longer term. Sometimes, consideration of the underlying hazard function associated with different models could lead to some models being rejected as implausible before attempting any statistical analysis.[75] For example, if it is believed that the underlying hazard function is non-monotonic then survival models with monotonic hazard functions such as a Gompertz distribution can be ruled out. Presentation of the empirical hazard function provides a useful descriptive summary of the sample data and further aids the identification of an appropriate model.[76] Despite the fact that the TSD guidance advocates that model selection should be based upon the hazard function, the shape of the empirical hazard function was only considered in around 40% of TAs.

Assuming proportional hazards is a modelling assumption that is convenient for generating a single estimate of treatment effect for drug registration purposes. Although an assumption of proportional hazards used in an HTA may be a consequence of using summary statistics generated from a meta-analysis, assuming proportional hazards when hazards are not proportional will generate biased estimates of population mean time-to-event.[77, 78] The underlying log hazards for each trial arm will not generally be parallel and proportional hazards models can often be ruled out. Despite the TSD guidance that proportional hazards assumption should be assessed, in some cases the assumption was applied without an assessment.

Statistical criteria can be used to assess absolute and relative goodness-of-fit, but a more important consideration is whether the models proposed are clinically plausible (both to the observed data as well as during the extrapolation phase). Despite the fact that the TSD recommends external validation using data and/or expert opinion, less than half of the TAs validated extrapolation approach according to clinical opinion and less than one third of the TAs used external data to choose the modelling approach

The apparent lack of consideration of these issues highlights the need for improvement in this area of assessment in NICE TAs. It is important to note that not all recommendations made in TSD 14 are relevant to all HTAs. For example, it is not necessary to assess whether hazards are proportional if there is no intention of making the proportional hazards assumption. In addition, for newer treatments with a novel mechanism of action, there may not be any relevant external sample data available with which to externally validate models. When these two issues (represented in table 1 as 1b and 2a) were removed from our criteria, 19% of the TAs reviewed followed the remaining suggestions made by TSD 14.

Several ERGs consistently use common parametric models, whereas it is notable that the LRiG ERG often use a hybrid approach based on an exponential survival function to extrapolate from a specified point in the tail of the Kaplan-Meier survival function. Whenever a piecewise model is used, uncertainty in the choice of change-point should be assessed.[79] Both of the modelling approaches are consistent with TSD 14, provided - as for any extrapolation - internal and external validity has been considered appropriately. However, companies are aware of disparities in the approaches taken by different ERGs, and this has been highlighted by companies involved in writing submissions to NICE, with a suggestion that this can be problematic.[80] However, in situations when extrapolation is required, it is important to consider alternative approaches to extrapolation, and the onus is on the company or ERG involved to justify the plausibility of the predictions.

In practice, it is possible that none of the commonly used parametric models described in TSD 14 will adequately represent expectations about the long-term hazards and those observed during the trial period. An observed hazard function, which is by definition an average over individuals, may exhibit odd shapes as a consequence of groups of patients with different underlying risk. For instance, strict entry criteria to clinical trials often result in a low short-term hazard function, which increases as individuals at higher risk experience an event, then decreases as a consequence of long-term survivors. In the longer term, the hazard might increase again if there are any extreme long term survivors because of age-related mortality. None of the commonly used survival models can represent these two turning points, and it is necessary to consider more complex models in these circumstances.

The focus of this manuscript was on compliance with TSD 14. TSD 14 was authored by the DSU which has a specific role in supporting the NICE TA programme. Other guides exist that provide recommendations for survival analysis and extrapolation within the context of HTA.[79, 81] The content of these existing guides is similar.

Since the publication of TSD 14, more complex survival models have been used in TAs, such as flexible parametric spline models and mixture models (including mixture cure models).[64, 66] There is some evidence to suggest that some of the more complex survival models are not well-understood by stakeholders in the NICE appraisal process. NICE TA483 was completed after the end date for our review but in the company submission, the ERG report and in the FAD a spline model with two knots was interpreted as implying that there were three heterogeneous subgroups of patients, each with a different survival profile that can be expressed as a combination of three survival functions.[82] In fact, this description is more akin to a mixture model, whereas a spline-based model simply represents a way of modelling a complex hazard function. That is, a spline-based model only describes the population (marginalised) hazard function, and cannot be used to make inferences about the unobserved hazard function for sub-groups of patients. The current guidance in TSD 14 for these more complex survival models is not detailed. Therefore, it may be useful to publish a new TSD to provide additional advice to companies, ERG and Appraisal Committee s on these more complex models.

Finally, TSD14 does not specify how parameters in survival models should be estimated and we did not extract this information during the review. However, it is likely that parameters are being estimated mostly using maximum likelihood, and that survival functions are generated using these estimates as plug-in values. Frequentist methods only consider sampling variation and do not deal with parameter uncertainty. Maximum likelihood estimates will generally correspond to the Bayesian posterior mode, whereas a Bayesian approach would estimate a survival function by taking the expectation with respect to the uncertain parameters. A maximum likelihood approach also requires an assumption of asymptotic multivariate normality in order to approximate joint posterior distributions with multivariate sampling distributions for use in probabilistic sensitivity analyses (which is inherently Bayesian). Although such approximations may be reasonable in some situations, it is precisely in situations when there are few events and limited follow-up with which to estimate parameters that the Bayesian approach and the incorporation of external information is important to represent genuine uncertainty. The review did not assess whether the survival model parameter uncertainty was correctly incorporated in the economic models as this would be difficult to assess without access to the economic models. Indeed, if there is external information used to validate extrapolations then it would be reasonable to include this in the analysis in order to strengthen inferences.

Conclusions

Extrapolation of hazard and survival functions is a complex area. However, making use of clinical knowledge of the disease process and mechanism of action of the treatment, and presenting the empirical hazard function should provide useful insights into plausible model selection. In general, our review has shown that while the majority of NICE cancer TAs assessed the absolute and relative goodness-of-fit of common parametric models to the sample data, other aspects such as consideration of the underlying hazard function and the use of external data have been given

relatively little attention. Overall, there is scope for improvement in the application of survival analysis methods used in NICE cancer technology appraisals to achieve greater transparency and consistency. An improvement could potentially be supported by explicit referencing in the NICE methods guide to TSD 14. Future research could investigate into the reasons for non-adherence to the TSD.

References

[1] Davies, C., Briggs, A., Lorgelly, P., Garellick, G. and Malchau, H., 2013. The "hazards" of extrapolating survival curves. Medical Decision Making, 33(3), pp.369-380.

[2] Kearns, B., Chilcott, J., Whyte, S., Preston, L. and Sadler, S., 2016. Cost-effectiveness of screening for ovarian cancer amongst postmenopausal women: a model-based economic evaluation. BMC medicine, 14(1), p.200.

[3] Latimer, N.R., 2013. Survival analysis for economic evaluations alongside clinical trials—extrapolation with patient-level data: inconsistencies, limitations, and a practical guide. Medical Decision Making, 33(6), pp.743-754.

[4] Guide to the methods of technology appraisal 2013: Process and methods, Published: 4 April 2013, nice.org.uk/process/pmg9

[5] Briggs A, Claxton K, Sculpher M. Decision modelling for health economic evaluation. Oxford University Press Inc., New York, 2006

[6] Sanders GD, Neumann PJ, Basu A, Brock DW, Feeny D, Krahn M. et al. Recommendations for Conduct, Methodological Practices, and Reporting of Cost-effectiveness Analyses: Second Panel on Cost-Effectiveness in Health and Medicine. JAMA 2016; 316(10):1093-1103

[7] Canadian Agency for Drugs and Technologies in Health, Guidelines for the economic evaluation of health technologies: Canada, 4th Edition, 2017

[8] Drummond MF, Schwartz JS, Jönsson B, Luce BR, Neumann PJ, Siebert U, Sullivan SD. Key principles for the improved conduct of health technology assessments for resource allocation decisions. Int J Technol Assess Health Care 2008;24(3):244-58

[9] NICE website: https://www.nice.org.uk/guidance/published?type=ta

[10] Latimer, N. NICE DSU Technical Support Document 14: Undertaking survival analysis for economic evaluations alongside clinical trials - extrapolation with patient-level data. 2011. http://scharr.dept.shef.ac.uk/nicedsu/wp-content/uploads/sites/7/2016/03/NICE-DSU-TSD-Survival-analysis.updated-March-2013.v2.pdf [Accessed July 2017]

[11] Kearns, B., Ara, R., Wailoo, A., Manca, A., Alava, M.H., Abrams, K. and Campbell, M., 2013. Good practice guidelines for the use of statistical regression models in economic evaluations. Pharmacoeconomics, 31(8), pp.643-652.

[12] Copyright 2017 Mendeley Ltd

[13] NICE TA250 Eribulin for the treatment of locally advanced or metastatic breast cancer (accessed July 2017) https://www.nice.org.uk/guidance/ta250

[14] NICE TA258 Erlotinib for the first-line treatment of locally advanced or metastatic EGFR-TK mutation-positive non-small-cell lung cancer (accessed July 2017) https://www.nice.org.uk/guidance/ta258

[15] NICE TA259 Abiraterone for castration-resistant metastatic prostate cancer previously treated with a docetaxel-containing regimen (accessed July 2017) https://www.nice.org.uk/guidance/ta259

[16] NICE TA263 Bevacizumab in combination with capecitabine for the first-line treatment of metastatic breast cancer (accessed July 2017) https://www.nice.org.uk/guidance/ta263

[17] NICE TA268 Ipilimumab for previously treated advanced (unresectable or metastatic) melanoma (accessed July 2017) https://www.nice.org.uk/guidance/ta268

[18] NICE TA269 Vemurafenib for treating locally advanced or metastatic BRAF V600 mutation-positive malignant melanoma (accessed July 2017) https://www.nice.org.uk/guidance/ta269

[19] NICE TA284 Bevacizumab in combination with paclitaxel and carboplatin for first-line treatment of advanced ovarian cancer (accessed July 2017) https://www.nice.org.uk/guidance/ta284

[20] NICE TA285 Bevacizumab in combination with gemcitabine and carboplatin for treating the first recurrence of platinum-sensitive advanced ovarian cancer (accessed July 2017) https://www.nice.org.uk/guidance/ta285

[21] NICE TA295 / TA421 Everolimus in combination with exemestane for treating advanced HER2-negative hormone-receptor-positive breast cancer after endocrine therapy (accessed July 2017) https://www.nice.org.uk/guidance/ta295

[22] NICE TA296 / TA422 Crizotinib for previously treated non-small-cell lung cancer associated with an anaplastic lymphoma kinase fusion gene (accessed July 2017) https://www.nice.org.uk/guidance/ta296

[23] NICE TA299 / TA401 Bosutinib for previously treated chronic myeloid leukaemia (accessed July 2017) https://www.nice.org.uk/guidance/ta299

[24] NICE TA306 Pixantrone monotherapy for treating multiply relapsed or refractory aggressive non-Hodgkin's B-cell lymphoma (accessed July 2017) https://www.nice.org.uk/guidance/ta306

[25] NICE TA307 Aflibercept in combination with irinotecan and fluorouracil-based therapy for treating metastatic colorectal cancer that has progressed following prior oxaliplatin-based chemotherapy (accessed July 2017) https://www.nice.org.uk/guidance/ta307

[26] NICE TA309 / TA402 Pemetrexed maintenance treatment following induction therapy with pemetrexed and cisplatin for non-squamous non-small-cell lung cancer (accessed July 2017) https://www.nice.org.uk/guidance/ta309

[27] NICE TA310 Afatinib for treating epidermal growth factor receptor mutation-positive locally advanced or metastatic non-small-cell lung cancer (accessed July 2017) https://www.nice.org.uk/guidance/ta310

[28] NICE TA311 Bortezomib for induction therapy in multiple myeloma before high-dose chemotherapy and autologous stem cell transplantation (accessed July 2017) https://www.nice.org.uk/guidance/ta311

[29] NICE TA316 Enzalutamide for metastatic hormone-relapsed prostate cancer previously treated with a docetaxel-containing regimen (accessed July 2017) https://www.nice.org.uk/guidance/ta316

[30] NICE TA319 Ipilimumab for previously untreated advanced (unresectable or metastatic) melanoma (accessed July 2017) https://www.nice.org.uk/guidance/ta319

[31] NICE TA321 Dabrafenib for treating unresectable or metastatic BRAF V600 mutation-positive melanoma (accessed July 2017) https://www.nice.org.uk/guidance/ta321

[32] NICE TA326 Imatinib for the adjuvant treatment of gastrointestinal stromal tumours (accessed July 2017) https://www.nice.org.uk/guidance/ta326

[33] NICE TA333 Axitinib for treating advanced renal cell carcinoma after failure of prior systemic treatment (accessed July 2017) https://www.nice.org.uk/guidance/ta333

[34] NICE TA338 Pomalidomide for relapsed and refractory multiple myeloma previously treated with lenalidomide and bortezomib (accessed July 2017) https://www.nice.org.uk/guidance/ta338

[35] NICE TA343 Obinutuzumab in combination with chlorambucil for untreated chronic lymphocytic leukaemia (accessed July 2017) https://www.nice.org.uk/guidance/ta343

[36] NICE TA344 Ofatumumab in combination with chlorambucil or bendamustine for untreated chronic lymphocytic leukaemia (accessed July 2017) https://www.nice.org.uk/guidance/ta344

[37] NICE TA347 Nintedanib for previously treated locally advanced, metastatic, or locally recurrent non-small-cell lung cancer (accessed July 2017) https://www.nice.org.uk/guidance/ta347

[38] NICE TA357 Pembrolizumab for treating advanced melanoma after disease progression with ipilimumab (accessed July 2017) https://www.nice.org.uk/guidance/ta357

[39] NICE TA359 Idelalisib for treating chronic lymphocytic leukaemia (accessed July 2017) https://www.nice.org.uk/guidance/ta359

[40] NICE TA360 Paclitaxel as albumin-bound nanoparticles in combination with gemcitabine for previously untreated metastatic pancreatic cancer (accessed July 2017) https://www.nice.org.uk/guidance/ta360

[41] NICE TA366 Pembrolizumab for advanced melanoma not previously treated with ipilimumab (accessed July 2017) https://www.nice.org.uk/guidance/ta366

[42] NICE TA370 Bortezomib for previously untreated mantle cell lymphoma (accessed July 2017)

https://www.nice.org.uk/guidance/ta370

[43] NICE TA371 / TA458 Trastuzumab emtansine for treating HER2-positive, unresectable locally advanced or metastatic breast cancer after treatment with trastuzumab and a taxane (accessed July 2017) https://www.nice.org.uk/guidance/ta371

[44] NICE TA376 / TA412 Radium-223 dichloride for treating hormone-relapsed prostate cancer with bone metastases https://www.nice.org.uk/guidance/ta376

[45] NICE TA377 Enzalutamide for treating metastatic hormone-relapsed prostate cancer before chemotherapy is indicated (accessed July 2017) https://www.nice.org.uk/guidance/ta377

[46] NICE TA378 Ramucirumab for treating advanced gastric cancer or gastro–oesophageal junction adenocarcinoma previously treated with chemotherapy (accessed July 2017) https://www.nice.org.uk/guidance/ta378

[47] NICE TA380 Panobinostat for treating multiple myeloma after at least 2 previous treatments (accessed July 2017) https://www.nice.org.uk/guidance/ta380

[48] NICE TA381 Olaparib for maintenance treatment of relapsed, platinum-sensitive, BRCA mutation-positive ovarian, fallopian tube and peritoneal cancer after response to second-line or subsequent platinum-based chemotherapy (accessed July 2017) https://www.nice.org.uk/guidance/ta381

[49] NICE TA384 Nivolumab for treating advanced (unresectable or metastatic) melanoma (accessed July 2017) https://www.nice.org.uk/guidance/ta384

[50] NICE TA386 Ruxolitinib for disease-related splenomegaly or symptoms in adults with myelofibrosis (accessed July 2017) https://www.nice.org.uk/guidance/ta386

[51] NICE TA387 Abiraterone for treating metastatic hormone-relapsed prostate cancer before chemotherapy is indicated (accessed July 2017) https://www.nice.org.uk/guidance/ta387

[52] NICE TA391 Cabazitaxel for hormone-relapsed metastatic prostate cancer treated with docetaxel (accessed July 2017) https://www.nice.org.uk/guidance/ta391

[53] NICE TA395 Ceritinib for the treatment of ALK positive nonsmall cell lung cancer previously treated with crizotinib (accessed July 2017) https://www.nice.org.uk/guidance/ta395

[54] NICE TA396 Trametinib in combination with dabrafenib for treating unresectable or metastatic melanoma (accessed July 2017) https://www.nice.org.uk/guidance/ta396

[55] NICE TA399 Azacitidine for treating acute myeloid leukaemia with more than 30% bone marrow blasts (accessed July 2017) https://www.nice.org.uk/guidance/ta399

[56] NICE TA400 Nivolumab in combination with ipilimumab for treating advanced melanoma (accessed July 2017) https://www.nice.org.uk/guidance/ta400

[57] NICE TA403 Ramucirumab for previously treated locally advanced or metastatic non-small-cell lung cancer (accessed July 2017) https://www.nice.org.uk/guidance/ta403

[58] NICE TA404 Degarelix for treating advanced hormone-dependent prostate cancer (accessed July 2017) https://www.nice.org.uk/guidance/ta404

[59] NICE TA405 Trifluridine–tipiracil for previously treated metastatic colorectal cancer (accessed July 2017) https://www.nice.org.uk/guidance/ta405

[60] NICE TA406 Crizotinib for untreated anaplastic lymphoma kinase-positive advanced non-small-cell lung cancer (accessed July 2017) https://www.nice.org.uk/guidance/ta406

[61] NICE TA408 Pegaspargase for treating acute lymphoblastic leukaemia (accessed July 2017) https://www.nice.org.uk/guidance/ta408

[62] NICE TA410 Talimogene laherparepvec for treating unresectable metastatic melanoma (accessed July 2017) https://www.nice.org.uk/guidance/ta410

[63] NICE TA411 Necitumumab for untreated advanced or metastatic squamous non-small-cell lung cancer (accessed July 2017) https://www.nice.org.uk/guidance/ta411

[64] NICE TA414 Cobimetinib in combination with vemurafenib for treating unresectable or metastatic BRAF V600 mutation-positive melanoma (accessed July 2017) https://www.nice.org.uk/guidance/ta414

[65] NICE TA416 Osimertinib for treating locally advanced or metastatic EGFR T790M mutation-positive non-small-cell lung cancer (accessed July 2017) https://www.nice.org.uk/guidance/ta416

[66] NICE TA417 Nivolumab for previously treated advanced renal cell carcinoma (accessed July 2017) https://www.nice.org.uk/guidance/ta417

[67] NICE TA424 Pertuzumab for the neoadjuvant treatment of HER2-positive breast cancer (accessed July 2017) https://www.nice.org.uk/guidance/ta424

[68] NICE TA428 Pembrolizumab for treating PD-L1-positive non-small-cell lung cancer after chemotherapy (accessed July 2017) https://www.nice.org.uk/guidance/ta428

[69] NICE TA429 Ibrutinib for previously treated chronic lymphocytic leukaemia and untreated chronic lymphocytic leukaemia with 17p deletion or TP53 mutation (accessed July 2017) https://www.nice.org.uk/guidance/ta429

[70] NICE TA432 Everolimus for the second-line treatment of metastatic renal cell carcinoma (accessed July 2017) https://www.nice.org.uk/guidance/ta432

[71] NICE TA440 Nanoliposomal irinotecan for the treatment of metastatic pancreatic cancer following gemcitabine-based therapy (accessed July 2017) https://www.nice.org.uk/guidance/ta440

[72] NICE TA374 Erlotinib and gefitinib for treating non-small-cell lung cancer that has progressed after prior chemotherapy (accessed July 2017) https://www.nice.org.uk/guidance/ta374

[73] Kearns, B., Jones, M.L., Stevenson, M. and Littlewood, C., 2013. Cabazitaxel for the second-line treatment of metastatic hormone-refractory prostate cancer: a NICE single technology appraisal. Pharmacoeconomics, 31(6), pp.479-488.

[74] Stevens, J.W., 2018. Using Evidence from Randomised Controlled Trials in Economic Models: What Information is Relevant and is There a Minimum Amount of Sample Data Required to Make Decisions?. PharmacoEconomics, 36(10), pp.1135-1141.

[75] Bullement, A., Latimer, N.R. and Gorrod, H.B., 2019. Survival Extrapolation in Cancer Immunotherapy: A Validation-Based Case Study. Value in Health, 22(3), pp.276-283.

[76] Casellas, J., 2007. Bayesian inference in a piecewise Weibull proportional hazards model with unknown change points. Journal of Animal Breeding and Genetics, 124(4), pp.176-184.

[77] Bagust, A. and Beale, S., 2014. Survival analysis and extrapolation modeling of time-to-event clinical trial data for economic evaluation: an alternative approach. Medical Decision Making, 34(3), pp.343-351.

[78] Guyot, P., Welton, N.J., Ouwens, M.J. and Ades, A.E., 2011. Survival time outcomes in randomized, controlled trials and meta-analyses: the parallel universes of efficacy and cost-effectiveness. Value in Health, 14(5), pp.640-646.

[79] Ishak, K.J., Kreif, N., Benedict, A. and Muszbek, N., 2013. Overview of parametric survival analysis for health-economic applications. Pharmacoeconomics, 31(8), pp.663-675.

[80] A. Benedict, N. Muszbek, 2014 “Survival Modelling in UK Oncology Technology Appraisals Since the Publication of Good Practice Guidelines” Evidera White paper https://www.evidera.com/resource/survival-modelling-in-uk-oncology-technology-appraisals-since-the-publication-of-good-practice-guidelines/ (accessed May 2018)

[81] Haines, P., Tremblay, G. and Briggs, A., 2015. A criterion-based approach for the systematic and transparent extrapolation of clinical trial survival data. Journal of Health Economics and Outcomes Research, 2(2), pp.147-160.

[82] NICE TA483 Nivolumab for previously treated squamous non-small-cell lung cancer (accessed May 2018)

Appendix Table A1: Extraction form section headings

Company submission OS/PFS
• Was an assessment made of the empirical hazard(s) based on the observed data and expert clinical input, including the log-cumulative hazard/quantile-quantile/other plots? [yes/no, free-text] • What distributions and/or statistical models were fitted to the observed data? [free-text] • Was an assessment made of the proportional hazards assumption? [yes/no] • How was the relative treatment effect modelled? [free-text] • Was an assessment made of the absolute goodness-of-fit i.e. interval validity? [yes/no, free-text] • Was an assessment made of the relative goodness-of-fit i.e. internal validity? [yes/no, free-text] • Was an assessment made of the external validity (i.e. clinical plausibility) of the extrapolations? [yes/no] • If, Yes, then what was this based on e.g. expert clinical opinion, external data? [free-text] • Was any external evidence used in the generation of the survivor functions? [free-text] • Was uncertainty accounted for during the observed and unobserved data periods? [yes/no, free-text] • Were summary statistics used to estimate survival for one (e.g. mean, median, proportion or a given time) or both arms (e.g. hazard ratio)? [yes/no, free-text] • What other considerations are relevant in the assessment (e.g. treatment switching)? [yes/no, free-text] • Did the modelling follow the recommendations of the DSU TSD? [yes/no] • If no, then in what ways did the modelling differ from the recommendations? [free-text]
ERG Report OS/PFS
• What criticisms did the ERG make of the company submission? [free-text] • What alternative approaches did they suggest and/or provide? [free-text]
FAD
• TA recommendation [free-text] • What issues relating to the extrapolation and modelling of the OS/PFS data did the Appraisal Committee mention in the FAD that affected the final decision? [free-text] • Which distribution and/or statistical model contributed to the Appraisal Committee s most plausible estimate of mean life years, mean QALYs and ICER? [free-text]