How do pharmaceutical companies model survival of cancer patients? A review of NICE single technology appraisals in 2017.

Daniel Gallacher[1], Peter Auguste, Martin Connock

Warwick Evidence, Warwick Medical School, University of Warwick, Coventry,
CV4 7AL, UK
1: ORCID: 0000-0003-0506-9384

Short title: **A review of survival extrapolations in NICE STAs.**

*Corresponding author:
Mr Daniel Gallacher
Research Fellow
Warwick Medical School
University of Warwick
Gibbet Hill Road
Coventry, UK, CV4 7AL
Tel: +44 (0)2476 574877
Email: d.gallacher@warwick.ac.uk

Abstract

Objectives: Before an intervention is publicly funded within the UK, the cost-effectiveness is assessed by the National Institute of Health and Care Excellence (NICE). The efficacy of an intervention across the patients' lifetime is often influential of the cost-effectiveness analyses, but is associated with large uncertainties. We reviewed committee documents containing company submissions and evidence review group (ERG) reports to establish the methods used when extrapolating survival data, whether these adhered to NICE Technical Support Document (TSD) 14, and how uncertainty was addressed.

Methods: A systematic search was completed on the NHS Evidence Search webpage limited to single technology appraisals of cancer interventions published in 2017, with information obtained from the NICE website.

Results: 28 appraisals were identified, covering 22 interventions across 18 diseases. Every economic model used parametric curves to model survival. All submissions used goodness-of-fit statistics and plausibility of extrapolations when selecting a parametric curve. 25 submissions considered alternate parametric curves in scenario analyses. Six submissions reported including the parameters of the survival curves in the probabilistic sensitivity analysis. ERGs agreed with the company's choice of parametric curve in nine appraisals, and agreed with all major survival-related assumptions in two appraisals.

Conclusions: TSD 14 on survival extrapolation was followed in all appraisals. Despite this, the choice of parametric curve remains subjective. Recent developments in Bayesian approaches to extrapolation are not implemented. More precise guidance on the selection of curves and modelling of uncertainty may reduce subjectivity, accelerating the appraisal process.

Abstract word count: 242

Main body word count (excluding tables and references): 2856

Key Words:

Technology Assessment, Biomedical

Survival Analysis

Models, Economic

Funding

This research received no specific grant from any funding agency, commercial or not-for-profit sectors.

Conflict of interest.

None of the authors have any conflicts of interest to declare.

Acknowledgements:

We thank Professor James Mason for his advice and support, and we are grateful to Rachel Court for her assistance with creating and implementing the search strategy.

1 Introduction

Globally, health service providers are under pressure to deliver a continuously improving standard of care with a finite pool of resources. This results in a desire for funding decision-makers, such as National Institute of Health and Care Excellence (NICE) in England, Wales and Northern Ireland, to obtain value for money, often setting thresholds where the price the decision-maker is willing to pay is based on the efficacy of the drug. This requires a common scale on which all diseases and interventions can be contrasted. In the UK, quality-adjusted life years (QALY) are commonly used, where patients' predicted life expectancy is multiplied by a value estimating the quality of their life across their lifetime. The resulting cost/QALY is considered alongside other factors, including whether an intervention can be considered to meet end-of-life criteria or an unmet need, before a decision on whether to reimburse a treatment in the National Health Service. Whilst unit prices are often fixed and contain little uncertainty, there is often considerable uncertainty over the clinical outcomes, such as life expectancy, and their associated resource use, with observed data often only spanning a couple of years. Commonly, progression-free survival (PFS) and overall survival (OS) data are modelled parametrically, with curves extrapolated across the patients' lifetime time horizon. The choice of parametric curves is often one of the most influential factors of a cost-effectiveness model, with subtle differences between parametric fits over a short follow-up period often yielding large disagreement upon extrapolation. Survival related assumptions often influence the manufacturers discount on the list price (e.g. through a commercial access agreement or patient access scheme) to ensure that the cost-effectiveness results fall within the thresholds NICE considers to be cost-effective. OS extrapolation is often more influential on a cost-effectiveness analysis, as PFS data are more mature with the PFS extrapolations containing less uncertainty. The NICE Decision Support Unit (DSU) has published Technical Support Document (TSD) 14 which suggests methods on how to assess the suitability of a survival model [1], though it remains unclear how closely this is adhered to. One of NICE's types of appraisal is the single technology appraisal (STA), where a single intervention is assessed for a single indication. The manufacturer is invited to submit evidence to support the

decision making process. This evidence is assessed by an interdisciplinary evidence review group (ERG), who produce an independent and unbiased report of the evidence that is presented alongside the manufacturer submission to a NICE appraisal committee. The committee then decide whether to fund the intervention, based on the clinical and cost-effectiveness evidence. Further information on the single technology process can be found on the NICE website [2].

This paper presents a review of STAs of cancer treatments whose guidance was published by NICE in 2017, focusing on the approaches to the extrapolation of survival data. It set out to achieve the following aims:

- Identify methods used by companies in their approaches to survival extrapolation, and investigate whether these adhere to NICE guidance
- Identify whether the methods used by company's result in robust selection of extrapolation and investigate how uncertainty in extrapolations is accounted for

To achieve these aims, we focused on appraisals of interventions targeting cancers to ensure relevant survival information would be present. We limited the search to STAs to maximise consistency in reporting standards.

1.1 Summary of NICE TSD 14 on Survival Extrapolation

NICE TSD 14 [1] is a report by NICE's Decision Support Unit, providing guidance on survival analysis and extrapolation of survival data within economic evaluations. Section 3 of TSD 14 lists some suggested methods for assessing the suitability of survival models. These are:

Visual Inspection: Comparing the parametric models to the observed Kaplan-Meier data.

Log-Cumulative Hazard Plots: TSD 14 suggests plotting the observed log-cumulative hazard data, usually against log-time, to assess the behaviour of the hazard from the observed data, which can assess the suitability of the exponential or Weibull functions. The Log-Cumulative hazard plot can also be used to assess the hazard proportionality of different treatment arms, helping to decide

whether an assumption of proportionality should be maintained in the extrapolation. Other plots such as Quantile-Quantile (QQ) and graphs of residuals (e.g. Martingale, Schoenfeld) can also be considered.

Information Criterion: The most commonly used information criteria are the Akaike information criterion (AIC) and Bayesian information criterion (BIC). These both allow a comparison of non-nested models, which is necessary due to the differing functional forms of the different parametric models.

Clinical validity and External Data: Comparison of extrapolations of parametric models to predictions made by expert clinical opinion and to data with longer follow-up of similar patients and treatments, typically from other trials or registry databases.

Unless separate forms are strongly justified with clinical expert opinion, biological plausibility and robust statistical analysis, the TSD recommends using the same parametric form for both the intervention and the comparator. An example of an appraisal where separate forms may have been appropriate was TA519 where pembrolizumab (an immunotherapy) was compared to chemotherapies in patients with urothelial cancer [3].

2 Methods

2.1 Study Eligibility Criteria

For inclusion in this review, only single technology appraisals of interventions for cancerous diseases published in 2017 were targeted.

2.1 Search Strategy

A systematic search was completed on the NHS Evidence Search webpage on 16/01/2018, using the terms neoplasm* or neoplasia or tumor* or tumour* or cancer* or carcinoma* or malignan* or adenocarcinoma* or metasta*, filtered to guidance published by NICE in the area of Drugs and

Technologies in 2017. A broader search of additional electronic databases was not required due to the particular focus of this review.

2.2 Search Selection

PA and DG independently assessed whether the studies met the inclusion criteria initially at title and abstract, and then at full text, with any disagreements resolved through discussion or recourse to a third reviewer (MC).

2.3 Data Extraction

Information from each appraisal were collected from the initial set of relevant committee papers into a data extraction form (available on request), performed by DG. Median and maximum follow-up durations were extracted from text where available, or from Kaplan-Meier plots. Where multiple trials were considered with a single appraisal, the trial with the longest follow-up was used for this analysis.

2.4 Quality Assessment

As all appraisals are scrutinised by an ERG and NICE technical team, it was not necessary to use further quality assessment tools.

3 Results

Figure 1 shows the outcomes of the search and screening process, including the reasons for exclusion, in a Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) diagram [4] .

The final 28 pieces of guidance covered 22 primary interventions across 16 different diseases. NICE recommended funding through the NHS for 18 of the appraisals [5-22], eight were recommended for funding through the CDF [23-30] and two were not recommended for funding [31, 32].

An overview of appraisals included in this review is presented in Table 1 and a summary of the findings is shown in Table 2.

Four submissions used a Markov multistate model to assess cost-effectiveness [13, 19, 20, 30], and 21 used a Partitioned Survival model [5-12, 14, 16-18, 21-28, 31, 32]. It was unclear for two submissions which modelling approach was implemented [15, 17], and for a further submission both a semi-Markov model and partitioned survival model were presented [29].

Every appraisal used frequentist parametric models to extrapolate at least one set of time-to-event data. In one appraisal splines were submitted as the company base case [25], however all others opted for a standard parametric form, either exponential, Weibull, log-normal, log-logistic, gamma, generalised gamma or Gompertz. Piecewise modelling featured in 11 appraisals [5, 6, 8, 10, 12, 14, 19, 22-25], usually featuring initial KM data being implemented into the economic model, followed by a parametric curve after a certain point in time. This approach was more commonly preferred by ERGs (8 appraisals) than companies (4 appraisals). Occasionally, a combination of distinct parametric curves were used to model the same survival outcome, e.g. when specific groups of patients had differing levels of risk, which was more common in the Markov models [7, 12, 15, 20, 29]. In one appraisal a company choose to model a survival outcome without extrapolating using a parametric form, and instead modelled observed KM data [28], compared with ERGs favouring this approach in three appraisals [16, 28, 32].

The time horizon of 27 models was reported, ranging from 10 to 100 years, with a mean of 31 years. All models used survival data from at least one clinical trial. The length of maximum observed follow-up periods of the main contributing trials, ranged from 1.4 years to 6.8 years, with a mean of 3.1 years. The percentage of each model's time horizon supported by observed data ranged from 5.8% to 37.5% with a mean of 12.4%. When the analysis was repeated using only reported median follow-up, the observed periods ranged from 0.4 years to 5.9 years, mean 2.0 years and corresponding observed time horizon percentages ranged from 2.0% to 11.6%, mean 6.5%.

All submissions reported using goodness of fit statistics (e.g. AIC and BIC) and plausibility of predictions when selecting the parametric curve. Eighteen reports also considered visual fit of parametric curves to the observed data, predominantly overlaid on a KM plot [8-11, 13, 18-23, 25-31]. Six submissions compared extrapolations to external sources of data [12, 17, 18, 24, 25, 28]. Five presented log-cumulative hazard plots [6, 10, 14, 21, 27], and two considered QQ-plots [6, 21]. Only two submissions reported verifying the consistency between PFS and OS curves [15, 26]. All appraisals used the same functional form for all modelled comparators.

Three company submissions did not present cost-effectiveness results for alternative parametric curves [5, 10, 14]. Five appraisals reported the inclusion of the survival distribution parameters in their probabilistic sensitivity analysis (PSA) [18, 19, 21, 29, 32] although it was common for reports to omit the full list of parameters included in the PSA, and it is possible that more PSAs did include survival parameters.

Each submission was critically appraised by an ERG who agreed with the company's choice of parametric fits in nine appraisals [8, 10, 14, 15, 18, 19, 21, 26, 31]. In two appraisals it was unclear whether the ERG maintained the company's choice or chose different curves [9, 16], meaning in most appraisals the ERG disagreed with the company's choice. In terms of broader survival assumptions (including choice of parametric curve), in three appraisals the ERG agreed with all major assumptions made by the pharmaceutical company [8, 18, 26], and in one of these cases the NICE committee preferred different survival assumptions to those agreed by the ERG and company [26]. In all other appraisals, the ERG disagreed with at least one major assumption of the survival extrapolation, for reasons including the methods used for adjusting for crossover, the consideration of internal and external data and the duration of treatment effect.

4 Discussion

This review of 28 NICE technology appraisal submissions showed that pharmaceutical companies generally take a transparent approach in their modelling and extrapolation of survival data in their

economic analyses. We found that extrapolation and curve selection were in line with TSD 14, with a range of parametric models use to predict future survival. We avoided making inference on the specific parametric curves chosen in each appraisal due the wide variety of cancers and interventions covered.

Frequently the reporting of methods of curve selection was not comprehensive with only a couple of the recommended methods by NICE TSD 14 presented in a company submission. It was unclear why other methods were not presented, perhaps due to their discordance with the presented methods, for brevity or the fact that they were never utilised.

NICE STAs can often go on for many months, requiring multiple committee meetings and generating numerous sets of committee documents for a single appraisal. Inconsistency in NICE's provision of these documents meant it was unclear at how many committee meetings an appraisal was discussed, and masked the identity of the major points of contention. In addition, key information was often redacted, e.g. life-year estimates and resulting ICERs, meaning it was difficult to ascertain precisely how influential the survival curves were, and prevented a more quantitative analysis. It was also unclear whether the same definition of median follow-up was used across each appraisal [33]. It is possible that either the company or ERG changed their position on survival assumptions based on additional arguments or data being put forward, which were not captured by this review. The conduct of future reviews would be enhanced with uniformity in the presentation of documentation by NICE, and greater consistency across companies in their presentation of survival extrapolation and justification.

The strengths of this review are that it captures the most up-to-date methods to extrapolate from survival data, across a wide range of interventions and cancers, as well as the types of economic model used. The systematic search identified all relevant appraisals, and summarised their methods of extrapolation and assessing uncertainty.

The limitations of this review are that it covers appraisals from a single year and is specific to cancer. This timeframe may mean it has missed other techniques used by companies.

The findings of this review are generaliseable to other disease areas where extrapolations of time-to-event outcomes are of primary concern to the decision-maker, particularly where partitioned survival models are suitable. They may not be representative of submissions to other decision makers who do not base decisions on patient population lifetime horizon.

One implication of this review is that the current version of TSD 14 may be insufficient in its instructions for the extrapolation of survival curves, due to the level of disagreement between companies and ERGs.

TSD 14 does acknowledge some limitations to its own recommendations, with an associated paper stating that the model selection algorithm "should not be viewed as a finished product" [34]. The limitations include that the AIC/BIC, visual fit and log-cumulative hazard plots only comment on the observed follow-up period, and do not necessarily infer on the suitability of the predictive ability of a model. These have also been discussed elsewhere [35, 36]. We agree with Bagust and Beale [35] in their response to TSD 14, and feel that assessing the visual fit against a KM plot and log cumulative hazard plots alone can be challenging as several parametric fits can often appear very similar, especially for late observations on the log-time scale. There are few alternative recommendations for robust extrapolations, for example Tremblay et al. (2015) [37] published five criteria for transparent extrapolation of survival data, though they are broadly similar to TSD 14. Their fourth criterion was to establish uncertainty in the estimate of marginal difference in treatment effects using bootstrapping, with high uncertainty indicating low robustness. The fifth criterion was to compare the similarity of the observed and extrapolated gains, using a rule of thumb that the ratio of relative difference in the extrapolated period divided by the number of months extrapolated (post-observed period) should not exceed the ratio of the incremental difference for the observed period over the duration of the observed period. These criteria are relatively untested, it is uncertain

whether they are suitable for discriminating between different parametric models or whether following them would recommend a different model to following TSD 14.

The TSD is critical of using -2 log-likelihood statistic, instead favouring the AIC and BIC both of which use this statistic in their calculation. Both AIC and BIC favour a parsimonious model, that is considering both the fit to the data and the number of parameters in the model. Whilst such an approach should be encouraged when adjusting for confounding effects, it is unclear whether using AIC/BIC over -2*log-likelihood statistic leads to a more accurate extrapolation when comparing different parametric forms which have similar numbers of parameters.

We would suggest also comparing a smoothed hazard plot against predicted hazards from the parametric models, without necessarily rescaling time on the log scale. This would allow for easier identification of deviation from the predicted hazards at the later stages of follow-up, and for easier distinction between the parametric curves. This should be done in consideration with the number of patients remaining at risk, as late events may incorrectly make a model appear to be a poor fit.

Another implication is that there is clear potential for companies to implement modern approaches into their economic models. Due to lengthy timelines of the appraisal process, it is possible that the most recent approaches to extrapolating and capturing uncertainty are unlikely to be seen in a STA. There may be resistance to change, with manufacturers perhaps wary of experimenting with new techniques when the approval of their intervention is at stake.

The appraisals in this review all used frequentist methodology to extrapolate survival data and explore uncertainty, despite alternative approaches being available. These approaches may fail to capture uncertainty or unusual hazard behaviour, and often exclude prior information. Poly- models, such as the Poly-Weibull model [38], assume the presence of multiple independent sources of risk that operate additively. The Poly-Weibull hazard function emerges as the sum of the independent Weibull hazards. Demiris et al found the Poly-Weibull gave a better fit compared to a single Weibull

model in their examples, however Poly-model fitting can be challenging in the absence of causes of event or when the multiple hazards do not appear distinctly within the modelled data [38].

Negrin et al [39] use Bayesian Model Averaging (BMA) to capture the uncertainty in the extrapolation of survival curves, which may be superior to conducting separate PSA and scenario analyses with the survival curves. Using BMA to combine the estimates of each parametric fit can produce a prediction for mean survival and a credible interval, removing the need to focus on a single parametric model. Jackson et al review methods of including external data into survival extrapolations [40], which may further improve their accuracy.

This review highlights the vast uncertainty that remains in many technology assessments, and raises the question of whether clinical trials should be designed with greater consideration of the funding decision-maker.

Further research is recommended into when survival data are mature enough to produce reliable extrapolations. It is plausible that before a certain length of follow-up and number of events, extrapolation is not reliable. This could reduce the need for adjustments to extrapolations, such as waning treatment effects, which are often applied when initial extrapolations are implausible but introduce additional uncertainty for the company, ERG and appraisal committee. Appraisals submitted prior without mature data could be made on a temporary basis, similar to the current running of the Cancer Drugs Fund (CDF) in the UK, where the CDF allows patients access to interventions, which have demonstrated potential to be cost-effectiveness, but where uncertainty remains.

5 Conclusion

Cancer STAs use frequentist parametric approaches to extrapolate survival and explore uncertainty. Despite adhering to TSD 14, pharmaceutical companies specify parametric curves and other assumptions that are routinely rejected by ERGs. More thorough guidance is recommended to

ensure methods of curve selection are consistent across appraisals. As extrapolations account for such a high proportion of an economic model's time horizon, it is critical that extrapolations are supported with strong justification. . Recent developments in Bayesian approaches to extrapolation and uncertainty were not implemented.

Author Contributions

DG generated the initial research idea and drafted the manuscript. DG and PA reviewed the eligibility of the submissions. DG extracted information from eligible submissions. All authors reviewed the final draft.

Table 1: Overview of Single Technology Appraisals included in this review.

ID	Intervention	Disease	Time Horizon	Max Observed follow up (months)	Analysis Type	Used Goodness of fit statistics	Assessed Visual Fit	Checked Plausibility	Checked Consistency with external data	Checked Consistency between OS and PFS curves	Used Log-cumulative hazard plots	Used QQ plots	Included Survival Curve Parameters in probabilistic sensitivity analysis	Other curves explored	ERG agreed with curves	ERG agreed with all major survival assumptions	Company Preferred Parametric Fit	ERG Preferred Parametric Fit
TA427 [21]	Pomalidomide	Myeloma	15 years	29	Partitioned Survival Model	Yes	Yes	Yes	No	No	Yes	Yes	Yes	Yes	Yes	No	PFS: Expo OS: Gen Gamma	PFS: Expo OS: Gen Gamma
TA428 [10]	Pembrolizumab	Non-Small Cell Lung Cancer	20 years	24	Partitioned Survival Model	Yes	Yes	Yes	No	No	Yes	No	Unclear	No	Yes	No	PFS: KM + Gen Gamma OS: KM + Expo	PFS: KM + Gen Gamma OS: KM + Expo
TA429 [19]	Ibrutinib	Chronic Lymphocytic Leukaemia	20 years	18	Markov Multistate Model	Yes	Yes	Yes	No	No	No	No	Yes	Yes	No	No	PFS: Weibull OS: Lognormal + Expo	PFS: Expo OS: Expo
TA432 [9]	Everolimus	Renal Cell Carcinoma	12 years	19.6	Partitioned Survival Analysis	Yes	Yes	Yes	No	No	No	No	Unclear	Yes	Unclear	Unclear	PFS: Weibull OS: Weibull	Unclear
TA440 [32]	Pegylated liposomal irinotecan hydrochloride trihydrate (nal iri)	Pancreatic Cancer	10 years	36	Partitioned Survival Model	Yes	Yes	Yes	No	No	No	No	Yes	Yes	No	No	PFS: Lognormal OS: Lognormal	Preferred KM data without any parametric fit
TA446 [22]	Brentuximab vedotin	Hodgkin's Lymphoma	A = 70 years B = 80 years	84	Partitioned Survival Model	Yes	Yes	Yes	No	No	No	No	Unclear	Yes	Unclear	Unclear	PFS: Lognormal OS: KM + Expo	Unclear

TA447 [14]	Pembrolizumab	Non-Small Cell Lung Cancer	20 years	19.7	Partitioned Survival Model	Yes	Yes	Yes	No	No	Yes	No	Yes	No	Unclear	No	PFS: KM + Weibull OS: KM + Expo	Unclear
TA450 [18]	Blinatumomab	Leukaemia	50 years	36	Partitioned Survival Model	Yes	Yes	Yes	Yes	No	No	No	Yes	Yes	Yes	Yes	PFS: Gen Gamma OS: Gompertz	PFS: Gen Gamma OS: Gompertz
TA451 [20]	Ponatinib	Leukaemia	100 years	NR	Markov Multistate Model	Yes	Yes	Yes	No	No	No	No	Unclear	Yes	No	No	PFS: Multiple OS: Multiple	PFS: Multiple OS: Multiple
TA457 [17]	Carfilzomib	Myeloma	40 years	53	Unclear	Yes	Yes	Yes	Yes	No	No	No	Unclear	Yes	No	No	PFS: Weibull OS: Weibull	PFS: Weibull OS: Gompertz
TA458 [8]	Trastuzumab Emtansine	Breast Cancer	15 years	27	Partitioned Survival Model	Yes	Yes	Yes	No	No	No	No	Unclear	Yes	Yes	Yes	PFS: KM + Gamma OS: Gamma	PFS: KM + Gamma OS: Gamma
TA462 [15]	Nivolumab	Hodgkin Lymphoma	40 years	28	Unclear	Yes	Yes	Yes	No	Yes	No	No	Unclear	Yes	Yes	No	PFS: Expo or Lognormal OS: Expo or Weibull	PFS: Expo or Lognormal OS: Expo or Weibull
TA463 [11]	Cabozantinib	Kidney Cancer	30 years	28.7	Partitioned Survival Model	Yes	Yes	Yes	No	No	No	No	Unclear	Yes	No	No	PFS: Loglogistic OS: Loglogistic	PFS: Loglogistic OS: Weibull
TA465 [28]	Olaratumab	Soft Tissue Sarcoma	25 years	47	Partitioned Survival Model	Yes	Yes	Yes	Yes	No	No	No	Yes	Yes	No	No	PFS: KM OS: Gen Gamma	PFS: KM OS: Lognormal
TA472 [29]	Obinutuzumab	Follicular Lymphoma	25 years	54	Company chose Semi	Yes	Yes	Yes	No	No	No	No	Unclear	No	No	No	PFS: Multiple	PFS: Multiple OS: Expo

					Markov Model. ERG preferred Partitioned Survival Model											OS: Multiple	
TA473 [5]	Cetuximab	Head/Neck Cancer	"lifetime"	60	Partitioned Survival Model	Yes	No	Yes	No	No	No	No	Yes	Yes	No	No	PFS: Weibull OS: Weibull	PFS: Expo OS: KM + Expo
TA476 [6]	Paclitaxel	Pancreatic Cancer	10 years	45	Partitioned Survival Model	Yes	Yes	Yes	No	No	Yes	Yes	Unclear	Yes	No	No	PFS: Gamma OS: Gamma	PFS: KM + Gamma OS: KM + Gamma
TA478 [16]	Brentuximab Vedotin	Lymphoma	60 years	82	Partitioned Survival Model	Yes	Yes	Yes	No	No	No	No	Yes	Yes	No	No	PFS: Lognormal OS: Gamma	PFS: KM OS: KM
TA483 [25]	Nivolumab	Non-Small Cell Lung Cancer	20 years	38	Partitioned Survival Model	Yes	Yes	Yes	Yes	No	No	No	Unclear	Yes	No	No	PFS: Spline OS: Loglogistic	PFS: KM + Expo OS: KM + Expo
TA484 [24]	Nivolumab	Non-Small Cell Lung Cancer	20 years	24	Partitioned Survival Model	Yes	Yes	Yes	Yes	No	No	No	Yes	No	No	No	PFS: Gen Gamma OS: Gen Gamma	PFS: KM + Expo OS: KM + Expo
TA487 [27]	Venetoclax	Leukaemia	20 years	24.7	Partitioned Survival Model	Yes	Yes	Yes	No	No	Yes	No	Yes	Yes	No	No	PFS: Weibull OS: Weibull	PFS: ERG Weibull OS: ERG Weibull
TA488 [7]	Regorafenib	Gastro-intestinal stromal tumours	40 years	45.9	Partitioned Survival Model	Yes	Yes	Yes	No	No	No	No	Unclear	No	No	No	PFS: Lognormal OS: Loglogistic	PFS: Lognormal OS: multiple

TA489 [31]	Vismodegib	Basal Cell Carcinoma	30 years	45	Partitioned Survival Model	Yes	Yes	Yes	No	No	No	No	Yes	Yes	Yes	No	PFS: Weibull OS: multiple	PFS: Weibull OS: multiple
TA490 [26]	Nivolumab	Head/Neck Cancer	20 years	17	Partitioned Survival Model	Yes	Yes	Yes	No	Yes	No	No	Unclear	Yes	Yes	Yes	PFS: Gen Gamma OS: Lognormal	PFS: Gen Gamma OS: Lognormal
TA491 [30]	Ibrutinib	Walden-stroms macroglobu-linaemia	30 years	30	Markov Multistate Model	Yes	Yes	Yes	No	No	No	No	Yes	Yes	Yes	No	PFS: Weibull OS: Expo	PFS: Weibull OS: Expo
TA492 [23]	Atezolizumab	Urothelial cancer	20 years	24.5	Partitioned Survival Model	Yes	Yes	Yes	No	No	No	No	Unclear	Yes	No	No	PFS: Gen Gamma OS: Gen Gamma	PFS: KM + Expo OS: KM + Weibull
TA495 [12]	Palbociclib	Breast Cancer	40 years	40.5	Partitioned Survival Model	Yes	No	Yes	Yes	No	No	No	Yes	No	No	No	PFS: Weibull OS: Weibull	PFS: KM + Expo OS: Weibull
TA496 [13]	Ribociclib	Breast Cancer	40 years	34.75	Markov Multistate Model	Yes	Yes	Yes	No	No	No	No	Yes	Yes	No	No	PFS: Expo OS: Unclear	PFS: Expo OS: Unclear
SUMMARY						28 (100%)	26 (93%)	28 (100%)	6 (21%)	2 (7%)	5 (18%)	2 (7%)	6 (21%)	25 (89%)	9 (32%)	2 (7%)		

* indicates max follow-up when median was not reported.
ERG - Evidence Review Group; QQ – Quantile Quantile; Expo – Exponential; KM – Kaplan Meier; OS – Overall Survival; PFS – Progression Free Survival.

Note: a semi-Markov model's transition probabilities are dependent on the time spent in a health state. A Partitioned Survival Model estimates the proportion in each health state directly from survival curves using mutually exclusive health states.

Table 2: Summary of additional findings of this review.

	Range	Mean	Median (IQR)
Time Horizon (years)	10 to 100	31.4	25 (20, 40)
Max Observed Follow-up (years)	1.4 to 6.8	3.1	2.9 (2.0, 3.8)
Proportion of Time Horizon Observed using Max	5.8% to 37.5%	12.4%	10.1% (8.2%, 15.0%)
Median Observed Follow-up (years)	0.4 to 5.9	2.0	1.3 (1.2, 2.1)
Proportion of Time Horizon Observed using Median	2.0% to 11.6%	6.5%	6.7% (4.0%, 8.8%)

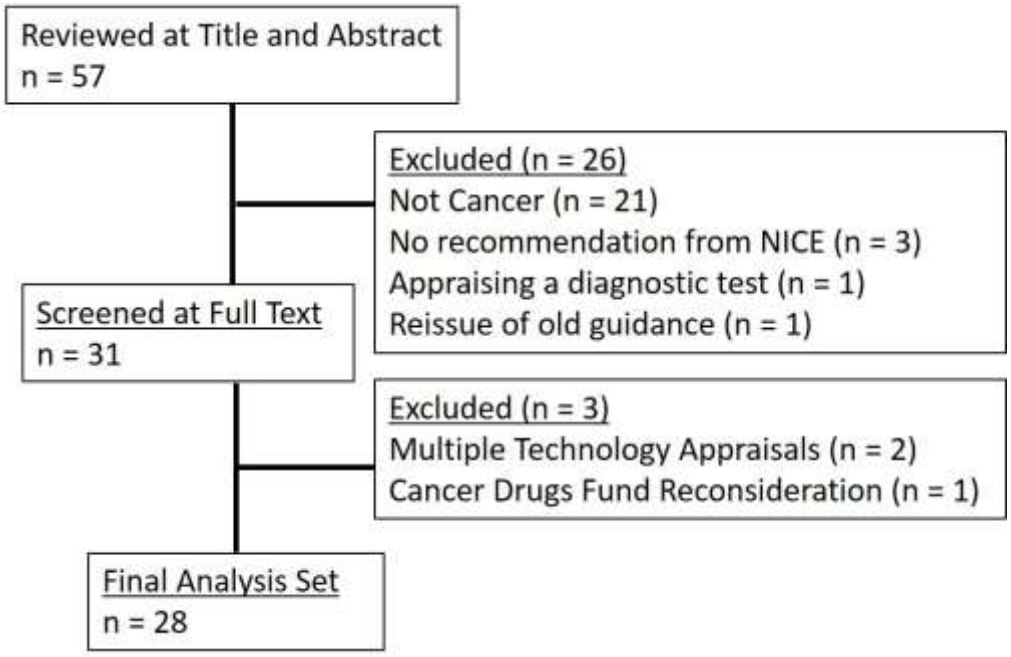

Figure 1: PRISMA flow diagram

References

1. Latimer N. NICE DSU Technical Support Document 14: Undertaking survival analysis for economic evaluations alongside clinical trials - extrapolation with patient-level data. 2011.
2. NICE Technology Appraisal Processes. [cited 2018 17 Oct]; Available from: https://www.nice.org.uk/about/what-we-do/our-programmes/nice-guidance/nice-technology-appraisal-guidance/process
3. Gallacher D, Armoiry X, Auguste P, Court R, Mantopoulos T, Patterson J, et al. Pembrolizumab for Previously Treated Advanced or Metastatic Urothelial Cancer: An Evidence Review Group Perspective of a NICE Single Technology Appraisal. PharmacoEconomics. 2018 July 21.
4. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JPA, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. 2009;339:b2700.
5. Cetuximab for treating recurrent or metastatic squamous cell cancer of the head and neck - guidance (TA473). 2017 2017/08/31.
6. Paclitaxel as albumin-bound nanoparticles with gemcitabine for untreated metastatic pancreatic cancer - guidance (TA476). 2017 2017/09/06.
7. Regorafenib for previously treated unresectable or metastatic gastrointestinal stromal tumours - guidance (TA488). 2017 2017/11/15.
8. Trastuzumab emtansine for treating HER2-positive advanced breast cancer after trastuzumab and a taxane - guidance (TA458). 2017 2017/07/19.
9. Everolimus for advanced renal cell carcinoma after previous treatment - guidance (TA432). 2017 2017/02/22.
10. Pembrolizumab for treating PD-L1-positive non-small-cell lung cancer after chemotherapy - guidance (TA428). 2017 2017/01/11.
11. Cabozantinib for previously treated advanced renal cell carcinoma - guidance (TA463). 2017 2017/08/09.
12. Palbociclib with an aromatase inhibitor for previously untreated, hormone receptor-positive, HER2-negative, locally advanced or metastatic breast cancer - guidance (TA495). 2017 2017/12/20.
13. Ribociclib with an aromatase inhibitor for previously untreated, hormone receptor-positive, HER2-negative, locally advanced or metastatic breast cancer - guidance (TA496). 2017 2017/12/20.
14. Pembrolizumab for untreated PD-L1-positive metastatic non-small-cell lung cancer – guidance (TA447). National Institute for Health and Care Excellence. 2017 2017/06/28.
15. Nivolumab for treating relapsed or refractory classical Hodgkin lymphoma - guidance (TA462). 2017 2017/07/26.
16. Brentuximab vedotin for treating relapsed or refractory systemic anaplastic large cell lymphoma - guidance (TA478). 2017 2017/10/04.
17. Carfilzomib for previously treated multiple myeloma - guidance (TA457). 2017 2017/07/19.
18. Blinatumomab for previously treated Philadelphia-chromosome-negative acute lymphoblastic leukaemia - guidance (TA450). 2017 2017/06/28.
19. Ibrutinib for previously treated chronic lymphocytic leukaemia and untreated chronic lymphocytic leukaemia with 17p deletion or TP53 mutation - guidance (TA429). 2017 2017/01/25.
20. Ponatinib for treating chronic myeloid leukaemia and acute lymphoblastic leukaemia - guidance (TA451). 2017 2017/06/28.
21. Pomalidomide for multiple myeloma previously treated with lenalidomide and bortezomib - guidance (TA427). 2017 2017/01/11.

22. Brentuximab vedotin for treating CD30-positive Hodgkin lymphoma - guidance (TA446). 2017 2017/06/28.
23. Atezolizumab for untreated PD-L1-positive locally advanced or metastatic urothelial cancer when cisplatin is unsuitable - guidance (TA492). 2017 2017/12/06.
24. Nivolumab for previously treated non-squamous non-small-cell lung cancer - guidance (TA484). 2017 2017/11/01.
25. Nivolumab for previously treated squamous non-small-cell lung cancer - guidance (TA483). 2017 2017/11/01.
26. Nivolumab for treating squamous cell carcinoma of the head and neck after platinum-based chemotherapy - guidance (TA490). 2017 2017/11/22.
27. Venetoclax for treating chronic lymphocytic leukaemia - guidance (TA487). 2017 2017/11/08.
28. Olaratumab in combination with doxorubicin for treating advanced soft tissue sarcoma - guidance (TA465). 2017 2017/08/09.
29. Obinutuzumab with bendamustine for treating follicular lymphoma refractory to rituximab - guidance (TA472). 2017 2017/08/30.
30. Ibrutinib for treating Waldenstrom's macroglobulinaemia - guidance (TA491). 2017 2017/11/22.
31. Vismodegib for treating basal cell carcinoma - guidance (TA489). 2017 2017/11/22.
32. Pegylated liposomal irinotecan for treating pancreatic cancer after gemcitabine - guidance (TA440). 2017 2017/04/26.
33. Betensky RA. Measures of Follow-up in Time-to-Event Studies: Why provide them and what should they be? Clinical trials (London, England). 2015 05/29;12(4):403-8.
34. Latimer NR. Survival Analysis for Economic Evaluations Alongside Clinical Trials—Extrapolation with Patient-Level Data:Inconsistencies, Limitations, and a Practical Guide. 2013;33(6):743-54.
35. Bagust A, Beale S. Survival Analysis and Extrapolation Modeling of Time-to-Event Clinical Trial Data for Economic Evaluation:An Alternative Approach. 2014;34(3):343-51.
36. Latimer NR. Response to “Survival Analysis and Extrapolation Modeling of Time-to-Event Clinical Trial Data for Economic Evaluation: An Alternative Approach” by Bagust and Beale. 2014;34(3):279-82.
37. Tremblay G, Haines P, Briggs A. A criterion-based approach for the systematic and transparent extrapolation of clinical trial survival data. Journal of Health Economics and Outcomes Research. 2015;2(2):147-60.
38. Demiris N, Lunn D, Sharples LD. Survival extrapolation using the poly-Weibull model. 2015;24(2):287-301.
39. Negrín MA, Nam J, Briggs AH. Bayesian Solutions for Handling Uncertainty in Survival Extrapolation. 2017;37(4):367-76.
40. Jackson C, Stevens J, Ren S, Latimer N, Bojke L, Manca A, et al. Extrapolating Survival from Randomized Trials Using External Data: A Review of Methods. 2017;37(4):377-90.