---

**Article:**
Tai, T.-A., Latimer, N. orcid.org/0000-0001-5304-5585, Benedict, A. et al. (2 more authors) (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 in Health, 24 (4). pp. 505-512. ISSN 1098-3015

https://doi.org/10.1016/j.jval.2020.10.016

---

**Abstract**

***Objectives:*** This research aims to explore how often the National Institute for Health and Care Excellence (NICE) uses immature overall survival (OS) data to inform reimbursement decisions on cancer treatments, and the implications of this for resource allocation decisions.

***Methods:*** NICE cancer technology appraisals (TAs) published between 2015 and 2017 were reviewed to determine the prevalence of using immature survival data. A case study was used to demonstrate the potential impact of basing decisions on immature data. The economic model submitted by the company was reconstructed and was populated first using survival data available at the time of the appraisal, and then using data from an updated data-cut published after the appraisal concluded. The incremental cost-effectiveness ratios (ICER) obtained using the different data-cuts were compared. Probabilistic sensitivity analysis (PSA) was undertaken and expected value of perfect information (EVPI) estimated.

***Results:*** 41% of NICE cancer TAs used immature data to inform reimbursement decisions. In the case study, NICE gave a positive recommendation for a limited patient subgroup, with ICERs too high in the complete patient population. ICERs were dramatically lower when the final data-cut was used, irrespective of the parametric model used to model survival. PSA and EVPI may not have fully characterised uncertainty as they did not account for structural uncertainty.

***Conclusion:*** Analyses of cancer treatments using immature survival data may result in incorrect estimates of survival benefit and cost-effectiveness, potentially leading to inappropriate funding decisions. This research highlights the importance of revisiting past decisions when updated data-cuts become available.

## Highlights

- Decisions on whether or not to reimburse new healthcare interventions are often made before clinical trial data are complete. For cancer treatments, extrapolation techniques are commonly used to estimate survival benefits beyond trial follow-up. However, extrapolation is associated with uncertainty, with different models or assumptions potentially leading to different results.
- The National Institute for Health and Care Excellence often base their decisions for cancer treatments on immature survival data. Data maturity can have a dramatic impact on cost-effectiveness conclusions, especially in cases where the shape of survival curves change significantly after the interim data-cut. Traditional probabilistic sensitivity analysis (PSA) only captures parameter uncertainty, and thus is unable to accurately characterise the uncertainty associated with data immaturity. This may result in inappropriate value of information estimates.
- It is critical for HTA authorities to revisit their decisions when important new data becomes available even when probabilistic sensitivity analyses suggest that the initial decision could be made with a high level of confidence, if there is an expectation that the PSA did not incorporate all types of uncertainty. Further research on methods for incorporating uncertainty around data immaturity into PSA is of vital importance.

## Introduction

Economic evaluation is an important part of Health Technology Assessment (HTA). Costs incurred and units of health gained are compared between interventions, allowing decision-makers to allocate finite resources[1]. The incremental cost-effectiveness ratio (ICER) represents the marginal cost per unit of health gained, and provides information on whether an intervention is 'cost-effective' with reference to a pre-specified threshold. In England, the National Institute for Health and Care Excellence (NICE) normally recommends a new intervention if its ICER is below £20,000 to £30,000 per quality-adjusted life year (QALY) gained. For treatments that meet end of life (EOL) criteria, QALY gains can be weighted by a factor of up to 1.7, implying a threshold of approximately £50,000 per QALY gained[2,3].

When treatments affect survival outcomes, economic evaluations need to adopt lifetime horizons[3-6]. Many regulatory bodies (such as the Food and Drug Administration (FDA) and European Medicines Agency (EMA)) have introduced accelerated or conditional approval schemes for licensing life-saving therapies[7-9] based on interim analysis or surrogate outcomes. To ensure timely patient access, HTA authorities aim to make decisions soon after drugs are licensed. Therefore, HTA decisions for cancer treatments often need to be based on clinical trials with incomplete, or "immature", survival data. For this reason, statistical models are used to extrapolate beyond the clinical trial follow-up. However, extrapolation introduces uncertainty and is challenging as different approaches can lead to different results[10,11]. Therefore, when substantial amounts of extrapolation are required, the reliability of cost-effectiveness conclusions is questionable.

Extrapolation methods represent a key focus in HTA submissions, due to their impact on cost-effectiveness results[12-17]. Discussions often centre on the plausibility of extrapolated portions of survival curves. However, a previous review shows that very few submissions to NICE thoroughly assess this[18]. Many researchers advocate the use of external information to guide extrapolation[10,18-20]. However, external data is likely only to be available for control treatments, which may not help guide estimates of long-term survival for new interventions[18], particularly for first-in-class therapies. Hence, funding decisions may be informed by unreliable cost-effectiveness analyses. This is important for HTA agencies around the world.

The objective of this research is to understand how often NICE makes decisions for cancer treatments using immature survival data. NICE is used as an example because their decisions are influential and are referenced internationally[21]. Dickson et al. 2018 found that an increasing number of NICE appraisals were based on trials with a single-arm, small sample size or limited follow-up, resulting in increased uncertainty which was often not appropriately parameterised in economic models, or sufficiently explored in scenario analyses[22]. We build upon this, and also present a case study of a first-in-class therapy to investigate the potential implications of using immature survival data for decision-making.

First we present our review of NICE cancer technology appraisals (TAs) with respect to survival data immaturity. Next, our case study is reported. Finally, we offer discussion, taking into account our review and case study.

**Review of NICE Cancer Technology Appraisals**

This review aimed to assess how often NICE bases reimbursement decisions on survival data considered to be immature, how Evidence Review Groups (ERGs) and Appraisal Committees (ACs) define survival data maturity, and how data maturity affects NICE's decisions.

### *Review Approach*

The review was conducted in July 2018. NICE single technology appraisals (STAs) published between January 2015 and December 2017 for cancer treatments were reviewed[23]. Information on the TA, clinical trial characteristics and data maturity was extracted from the company submissions (CS), ERG reports and final guidance documents. The level of survival data maturity was categorised as 'mature', 'immature', 'partial information' and 'not mentioned' based on statements made by ERGs and/or ACs. This permitted an understanding of what ERGs/ACs considered to represent 'mature' data. Statements were identified by searching for the terms 'mature', 'immature', 'maturity' and 'immaturity' in the ERG reports and final guidance. TAs identified as providing 'partial information' on data maturity were those involving 2 or more clinical trials, not all of which were commented on by the ERG or AC with respect to survival data maturity. 'Not mentioned' TAs were those where ERGs and ACs made no statement around survival data maturity.

### *Findings*

Forty-nine cancer STAs were reviewed (Table S.1, supplementary material).

*Definition of survival data maturity by ERGs or ACs.* Statements on data maturity were primarily driven by the proportion of deaths in the pivotal trials[24-28]. In STAs that provided

information on the proportion of deaths, the majority of 'mature' and 'immature' cases had proportions of deaths over 70% and under 50% respectively (Figure 1). Uncertainty in the trend of survival curves was also considered. For instance, in TA450 the pivotal trial was referred to as 'immature' even though 62% of patients had died, seemingly because overall survival (OS) curves plateaued and converged in the long-term when data were sparse due to few patients remaining at risk[29].

*Prevalence of using immature survival data in NICE decisions.* 41% (n=20) of reviewed STAs were considered by ERGs/ACs to involve immature survival data; 20% (n=10) were considered mature; 2% (n=1) had partial information and in 37% (n=18) data maturity was not commented on (Table S.1, supplementary material). Hence, NICE frequently based recommendations for cancer treatments on data considered by ERGs/ACs to be immature.

*Characteristics of 'mature' and 'immature' STAs.* Clinical trials considered to provide immature data were more likely to be single-arm and early phase trials (i.e. phase I or II), or those presenting interim analyses. 75% (n=15) of 'immature' STAs versus 10% (n=1) 'mature' STA used interim trial results. 25% (n=5) 'immature' STAs versus 10% (n=1) 'mature' STA involved a single-arm trial design. 40% (n=8) of 'immature' STAs versus 10% (n=1) 'mature' STA employed early-phase trials (Table S.2, supplementary material). This implies that many 'immature' TAs may suffer from uncertainty associated not only with survival data but also small sample size and a need for making indirect comparisons[30].

*Recommendations made by NICE.* When preferred ICERs fell on the borderline or out of the range normally considered cost-effective, 'mature' STAs were more likely to be given negative recommendations while 'immature' STAs were recommended within the Cancer Drugs Fund (CDF)[28,31,32] (Table S.3, supplementary materials). 50% of 'mature' STAs (n=5) received

negative recommendations; while 35% of 'immature' STAs (n = 7) resulted in CDF recommendations and none resulted in negative decisions. The positive decisions for 'immature' STAs (65%, n = 13) were sometimes 'restricted'. For example, in TA381[33], the recommendation was positive for a sub-group, due to the ICER being too high in the broader population.

**Case Study**

TA381[33] was selected as the case study to investigate the potential implications of making funding decisions based on immature survival data. Survival data was considered to be immature by the ERG/AC, a data-cut with relatively complete survival data was published after the NICE appraisal, and data available was sufficient to allow the economic model used in the STA to be replicated. TA381 assessed olaparib for relapsed, platinum-sensitive, BReast CAncer mutation-positive (BRCAm) ovarian, fallopian tube and peritoneal cancer after response to platinum-based chemotherapy. The Phase II pivotal trial (Study 19), a double-blind, randomised and placebo-controlled trial, enrolled patients who had the BRCA mutation and previously had two or more lines of chemotherapy (2L+ BRCAm)[34]. The company submitted a semi-Markov economic model, and the ERG built a partitioned survival model (PartSM) to validate the company's results. It was concluded that treatment in the 2L+BRCAm group was not cost-effective, but a positive recommendation was made for a smaller subgroup (third-line patients – 3L+ BRCAm), based upon the PartSM model.

Our case study focuses on the 2L+ BRCAm population. A November 2012 data-cut from Study 19, with 3-year follow-up, was used in TA381. Subsequently two more data-cuts were published, with median follow-up of approximately 6 and 6.5 years[35,36]. The data available for

the three main clinical endpoints in the various data-cuts are summarised in Table 1. The 'original data-cut' refers to the 2012 data-cut. The 'latest data-cut' refers to the data-cut with the most mature data for each endpoint, which differs for different endpoints.

***Methods***

First, pseudo individual patient data (IPD) were reconstructed for the original data-cut, parametric survival models were fitted and a preferred model selected. Second, the economic model used in TA381 was recreated. Third, the same process was performed for the latest data-cut. Finally, cost-effectiveness results were compared between different data-cuts.

*Survival Modelling.* Pseudo IPD were reconstructed from published KM curves using the algorithm developed by Guyot et al.[37]. Survival model selection was based on NICE Decision Support Unit Technical Support Document (TSD) 14[10,18] – reflecting the approach used in TA381. All standard parametric distributions were fitted[38]. When these did not provide a good fit, flexible spline-based models were employed[39]. Model fit to the observed data was assessed by visual inspection, empirical hazards, Akaike's Information Criterion (AIC) and the Bayesian Information Criterion (BIC). To avoid making restrictive assumptions about the treatment effect survival curves were fit to each treatment arm independently, in line with the ERG's approach in TA381[40]. The extrapolated portion of the survival curves was assessed using external information including national cancer mortality statistics and comments made by the company and/or the ERG during TA381[41,42].

At the time of the original data-cut for Study 19, 23% of placebo group patients had received subsequent poly (adenosine diphosphate ribose) polymerase (PARP) inhibitors after disease progression or discontinuation of treatment[43]. In TA381, the company adjusted for this by

excluding trial sites where at least 1 patient had switched to PARP inhibitors, improving the hazard ratio (HR) for olaparib[43]. Adjustment analyses were not reported for the latest data-cut. Without IPD, treatment switching present in the latest data-cut could not be adjusted for so, for consistency, switching was not adjusted for in any of our analyses. Also, in TA381, survival models were fitted including prognostic baseline covariates. Without IPD, baseline covariates could not be included in our analyses.

*Construction of Economic Model.* The PartSM[44] model used in TA381 was replicated, including three main health states - progression-free, progressed disease and death (Figure S.12, supplementary material). In TA381 time to first subsequent therapy (TFST) was used as a proxy for progression-free survival (PFS), because PFS data had not been updated since June 2010. Model settings and input parameters followed those reported in TA381[45] – including a 15-year (lifetime) time horizon, a one-month cycle length and half-cycle correction. Model inputs are summarised in Table S.10 (supplementary material). The cost of olaparib was based on the list price as this information was redacted from the CS. A patient access scheme agreed during the TA was incorporated, whereby the company pays for olaparib for patients who remain on treatment after 15 months. A 3.5% discount rate for costs and QALYs was used, and the analysis was undertaken from the National Health Service (NHS) and Personal Social Service (PSS) perspective, according to the NICE reference case[3]. The reconstruction of the model was validated by comparing results to those reported in TA381 (see supplementary material).

*Model Analyses.* A threshold of £30,000 per QALY gained was used to examine whether NICE's decision may have differed based upon the latest data-cut. An implied EOL threshold of £50,000 was also used to assess the potential decision had EOL criteria been deemed to have been met[2]. Scenario analyses were performed to assess the impact of the choice of survival

models on the results. Parameter uncertainty was characterised using probabilistic sensitivity analysis (PSA). Distributions assigned to parameters followed those reported in the CS (Table S.10, supplementary material). Parameters of survival models were sampled from multivariate normal distributions using Cholesky decomposition[4]. 5,000 Monte Carlo simulations were run for PSA and results were illustrated using a cost-effectiveness plane (CEP), cost-effectiveness acceptability curves (CEAC) and frontiers (CEAF). Expected value of perfect information (EVPI) was derived using the Sheffield Accelerated Value of Information (SAVI) tool[46,47], withthe decision relevance horizon set to 5 years and the annual incidence of the target population estimated to be 450 patients, according to the CS[45].

*Software.* KM curves were digitized using WebPlotDigitizer[48]. Survival analysis was performed in R[49] using the flexsurv package[50]. The economic model was constructed using Microsoft Excel.

***Results***

*Survival Modelling and Replication of Economic Model.* The reconstructed IPD appeared to be a good representation of the trial data (Table S.5, supplementary material). A flexible spline model was used to model TFST and OS for the latest data-cut due to two turning points in the observed hazard functions. Model selection details are described in Table S.6 (supplementary material). The economic model used in TA381 was accurately replicated with costs, QALYs and ICERs very similar to those reported in TA381 for the equivalent analysis (Table S.8, supplementary material).

*Comparison of extrapolated and actual OS curves.* Figure 2 compares the OS curves projected using the immature data and the actual longer-term curves. There is a substantial discrepancy

between these curves for olaparib. The curves projected for the two treatment arms based on the original data-cut converged, while the subsequently observed KM curves remained clearly separated. Therefore, the curves projected based upon the immature survival data underestimated the longer-term survival benefit associated with olaparib, according to the subsequently observed longer-term data.

*Cost-Effectiveness Modelling with The Original and Latest Data-cut.* The analysis based on the later data-cut yielded similar costs to those estimated using the original data-cut, but estimated QALY gains increased significantly – from 0.36 to 0.80 QALYs. As a result, the ICER reduced dramatically from £101,467 to £45,787 (Table 2). The CEAC indicated a much higher probability of olaparib being cost-effective based on the latest data-cut (13% and 56% at thresholds of £30,000 and £50,000 per QALY gained respectively, Vs 2% and 18% based on the original data-cut) (Figure S.13, supplementary material). However, the incremental cost, QALY and net monetary benefit estimates associated with the analysis of the latest data-cut lie within the 95% credible intervals produced by the analysis of the original data-cut (Table 2).

At a threshold of £30,000 per QALY gained, the population EVPI was £152,141 and £2,201,146 for the original and latest data-cut respectively. The maximum EVPI across all thresholds was higher based upon the original data-cut, but was reached only at very high cost-effectiveness thresholds (around £100,000 per QALY) (see supplementary material). Based upon the latest data-cut, the maximum EVPI decreased – reflecting reduced uncertainty due to the additional data collected – but was reached at cost-effectiveness thresholds close to those used by NICE. Thus, at a threshold of £30,000 per QALY gained, the EVPI associated with the latest data-cut was actually higher than that associated with the original data-cut.

Scenario analyses for the original data-cut revealed that the ICER was sensitive to the choice of parametric model for OS – ranging between £61,684 and £280,487 per QALY gained depending on the model used. The ICER was much less sensitive to the choice of OS model for the latest data-cut – ranging from £33,004 to £43,195 per QALY gained (Table 3).

## Discussion

Our study reveals that NICE decisions for cancer therapies often have to be made based on immature survival data, which can have important implications for cost-effectiveness estimates. 41% of cancer STAs relied on immature survival data between 2015 and 2017. Our case study demonstrated that relying on immature survival data as the basis for model fitting may result in seriously inaccurate estimates of survival benefits. In our example, using more complete data halved the ICER compared to using the original immature data-cut. The original NICE recommendation was consistent with the evidence available at the time of the appraisal – but the recommendation may have been different if mature data had been available at the time of the decision.

The dramatic change in ICERs observed in our case study are due to a change in shape of the OS curve for olaparib after the original data-cut. KM curves for the two treatment arms converged shortly before the original data-cut but separated again afterward. Parametric models fit to the original data-cut were unable to predict this change. The sensitivity of the ICER at the original data-cut was tested using many different survival models for extrapolation, resulting in vastly diverging ICERs, but none were as low as that produced using longer-term data. Notably, scenario analyses using different survival models resulted in a much narrower range of ICERs when models were fitted to the later data-cut. This suggests that decision-makers should be

cautious about using immature data, provides rationale for re-visiting recommendations when more evidence becomes available, and suggests that there may be benefits associated with delaying reimbursement decisions until more mature survival data are available.

However, there are opportunity costs to delaying reimbursement decisions. Adopting a therapy that is not cost-effective is costly, but waiting for more complete data may result in QALY losses in the interim period if a treatment is later found to be cost-effective. This trade-off is of particular concern for life-threatening diseases, where there is a strong desire for rapid access to treatments. The CDF may be argued to strike a balance, allowing early patient access while collecting more evidence to reduce uncertainty[51]. In our case study, the survival benefit associated with olaparib appeared to be underestimated using immature data. In other cases, the reverse may be true. Of the 65% of STAs that resulted in positive recommendations based upon immature survival data, we do not know which interventions truly resulted in the survival benefits predicted during their Appraisal.

Our review shows that cancer drugs were never rejected when survival data were immature – 35% were placed in the CDF and 65% received full or restricted positive recommendations, while 50% of STAs involving mature data resulted in rejections. Drugs placed in the CDF are usually scheduled for re-assessment 2 years after the initial recommendation. For other drugs, NICE usually schedules re-assessment after 3 years. However, re-assessments for non-CDF drugs do not always involve a re-analysis of cost-effectiveness. The criteria that trigger a full re-assessment of a NICE appraisal are not entirely clear[52]. Our findings suggest that re-assessments of cost-effectiveness can be important even when sensitivity analyses suggest that initial decisions can be made with high levels of confidence. However, our analysis represents a single case-study – it is unclear whether re-assessment would result in such markedly different cost-

effectiveness estimates in other cases. Further case studies would be of value, as would research into methods for identifying cases in which re-analysis is likely to be most important.

In our case study, PSA results using the original data-cut suggested little decision uncertainty, with only 2% probability of olaparib being cost-effective and a low EVPI estimate. Due to the additional survival benefit shown at the latest data-cut, the probability of olaparib being cost-effective increased, resulting in an increased EVPI at a cost-effectiveness threshold of £30,000 per QALY gained. Given that cost-effectiveness estimates substantially changed, the original, low EVPI estimate may appear misleading. Traditional PSA only characterises parameter uncertainty, and therefore our original EVPI estimate did not incorporate structural uncertainty, or any additional uncertainty associated with immature survival data – therefore results may have appeared more certain than they actually were. In our case study, the incremental QALY benefit associated with olaparib according to the analysis of the latest data-cut lies within the credible interval produced by the analysis of the original data-cut, meaning that we cannot conclude definitively that parameter uncertainty inadequately characterised the uncertainty around the effectiveness of olaparib. However, it is reasonable to suggest that uncertainty may be more appropriately characterized and EVPI estimates may more accurately reflect the value of further research if all sources of uncertainty could be parameterised into PSA..

Our findings reflect those of Dickson et al. 2018, who found that NICE appraisals were often based on trials with limited follow-up, and that uncertainty was not appropriately parametrised in economic models[22]. Novel methods for parameterising uncertainty around survival extrapolation have been proposed[20,53-55]. Model averaging may be useful for incorporating uncertainty around the selection of survival models – although if no models provide good projections of future survival, averaging may not help. Mahon proposed the use of model averaging and a ‘temporal

uncertainty' parameter to quantify uncertainty around survival data in PSA[54]. These methods are not yet in common practice and require validation. Further research on methods that could incorporate uncertainty around survival data immaturity into CEA is of great importance.

Bayesian analyses may be helpful when reasonable expectations about long-term survival can be formed at the time of early data-cuts. External data and expert elicitation could be employed, using a Bayesian framework. For example, Guyot and colleagues used Bayesian multi-parameter evidence synthesis to incorporate external information into a survival model and reported improved extrapolation performance[20].If, at the time of the original Study 19 data-cut, there had been reason to expect a longer-term survival advantage associated with olaparib, this could have been parameterised into the model, which may have altered results. However, to the best of our knowledge, such information was not available at the time of the original data-cut.

In trials with small sample sizes, KM curves may have unexpected kinks due to chance – especially when numbers at risk become small. In our case study, it is important to consider whether the convergence of the olaparib and control group KM curves observed in the original data-cut or the subsequent divergence of these curves in the later data-cut was due to chance, or indeed whether there is a clinical and biologically plausible argument for why the curves may converge and then subsequently diverge. An ongoing phase III trial of olaparib in the same patient population[56] may provide useful information on this. Bagust et al. discuss why extrapolations based on short-term and long-term data may be different[57] suggesting that early censoring might be the cause. For the final analysis of a clinical trial patients are contacted to confirm their vital status, whereas this may not be required prior to an interim analysis. This can lead to early censoring and a lack of long-term information in interim datasets.

We are not aware of previous studies in the oncology field that have published updated cost-effectiveness results following the publication of updated data-cuts from clinical trials. Davies and colleagues provide an example outside oncology, comparing alternative hip replacement prosthesis[58]. The analysis resulted in contrasting cost-effectiveness conclusions when based upon 16 years of follow-up data compared to 8 years, again showing that analyses based on early data-cuts may be misleading.

A limitation of our review of NICE appraisals is that classifications of data maturity were based purely on statements made by the ERGs/ACs. These statements may be arbitrary, and in some cases maturity may have been referred to using terms not captured by our review. However, the review provides useful information on how data maturity has been considered by ERGs and appraisal committees, which was our aim. In addition, our review only included cancer TAs published between January 2015 and December 2017. It is unclear whether trends in the use of immature survival data have changed over time, and it is possible that the use of immature survival data in NICE appraisals has changed since 2018 - though we believe this is unlikely.

A key limitation of our case study is that the new Study 19 data-cut – and therefore our updated cost-effectiveness analysis – was not adjusted for treatment switching. Adjustment would be expected to further reduce the ICER, exacerbating the differences between the analyses based on the original and latest data-cuts. A further limitation is that we only investigated one case study and thus the generalisability of our findings is uncertain. Increased data maturity may not always drastically change cost-effectiveness results. However, knowing that this *can* be the case – based on the case study presented here – suggests that immature data should be used with extreme caution**.** Finally, full uncertainty was not parameterised in our PSA, due to our use of

traditional PSA methods, only including parameter uncertainty. This is likely to have contributed to EVPI under-estimates.

## Conclusion

NICE often needs to use immature survival data to make recommendations for cancer treatments. Analyses based on interim and longer-term survival data can result in very different cost-effectiveness estimates, as demonstrated by our case study. Uncertainty associated with analyses based on early data-cuts may not be appropriately characterised by PSA, resulting in under-estimates of the value of further research, because traditional PSA only captures parameter uncertainty. We advocate routine review of past decisions when updated data-cuts become available. Further research on methods for incorporating uncertainty associated with survival data immaturity into cost-effectiveness analysis is warranted.

## References


1. Drummond M, Sculpher MJ, Claxton K, Stoddart GL, Torrance GW. *Methods for the economic evaluation of health care programmes.* Fourth ed. Oxford, United Kingdom: Oxford University Press; 2015.

2. Paulden M. Recent amendments to NICE's value-based assessment of health technologies: implicitly inequitable? *Expert Rev Pharmacoecon Outcomes Res.* 2017;17(3):239-242.

3. National Institute for Health and Care Excellence (NICE). Guide to the methods of technology appraisal. https://www.nice.org.uk/process/pmg9/chapter/the-reference-case. Published 2013. Accessed August 16, 2018.

4. Briggs AH, Claxton K, Sculpher MJ. *Decision modelling for health economic evaluation.* Oxford: Oxford University Press; 2006.

5. Canadian Agency for Drugs and Technologies in Health (CADTH). Guidelines for the Economic Evaluation of Health Technologies: Canada. https://www.cadth.ca/sites/default/files/pdf/guidelines_for_the_economic_evaluation_of_health_technologies_canada_4th_ed.pdf. Published 2017. Accessed August 16, 2018.

6. Commonwealth of Australia. Guidelines for preparing a submission to the Pharmaceutical Benefits Advisory Committee. https://pbac.pbs.gov.au/content/information/files/pbac-guidelines-version-5.pdf. Published 2016. Accessed August 16, 2018.

7. Dagher R, Johnson J, Williams G, Keegan P, Pazdur R. Accelerated approval of oncology products: a decade of experience. *J Natl Cancer Inst.* 2004;96(20):1500-1509.

8. U.S. Food and Drug Administration (FDA). Fast Track, Breakthrough Therapy, Accelerated Approval, Priority Review. https://www.fda.gov/ForPatients/Approvals/Fast/default.htm. Accessed June 20, 2018.

9. The European Medicines Agency (EMA). Conditional marketing authorisation. http://www.ema.europa.eu/ema/index.jsp?curl=pages/regulation/general/general_content_000925.jsp. Accessed June 20, 2018.

10. Latimer N. Survival analysis for economic evaluations alongside clinical trials--extrapolation with patient-level data: inconsistencies, limitations, and a practical guide. *Med Decis Making.* 2013;33(6):743-754.

11. Gerdtham UG, Zethraeus N. Predicting survival in cost-effectiveness analyses based on clinical trials. *Int J Technol Assess Health Care.* 2003;19(3):507-512.

12. Grieve R, Hawkins N, Pennington M. Extrapolation of survival data in cost-effectiveness analyses: improving the current state of play. *Med Decis Making.* 2013;33(6):740-742.

13. Demiris N, Lunn D, Sharples LD. Survival extrapolation using the poly-Weibull model. *Stat Methods Med Res.* 2015;24(2):287-301.

14. Jackson CH, Sharples LD, Thompson SG. Survival models in health economic evaluations: balancing fit and parsimony to improve prediction. *Int J Biostat.* 2010;6(1):Article 34.

15. Demiris N, Sharples LD. Bayesian evidence synthesis to extrapolate survival estimates in cost-effectiveness studies. *Stat Med.* 2006;25(11):1960-1975.

16. Neyt M, Van Brabandt H, Devriese S, Van De Sande S. A cost-utility analysis of transcatheter aortic valve implantation in Belgium: focusing on a well-defined and identifiable population. *BMJ Open.* 2012;2(3).

17. Fleeman N, Bagust A, Richardson M, et al. Pembrolizumab for treating unresectable, metastatic melanoma after progression with ipilimumab. Produced by Liverpool Reviews and Implementation Group, commissioned by NIHR HTA Programme. https://www.nice.org.uk/guidance/ta357/documents/melanoma-unresectable-metastatic-pembrolizumab-after-ipilimumab-id760-committee-papers2. Published July 2015. Accessed July 4, 2018.

18. Latimer N. NICE DSU Technical Support Document 14. Undertaking survival analysis for economic evaluations alongside clinical trials - extrapolation with patient-level data. http://www.nicedsu.org.uk/. Published 2011. Accessed August 16, 2018.

19. Jackson C, Stevens J, Ren S, et al. Extrapolating Survival from Randomized Trials Using External Data: A Review of Methods. *Med Decis Making.* 2017;37(4):377-390.

20. Guyot P, Ades AE, Beasley M, Lueza B, Pignon JP, Welton NJ. Extrapolation of Survival Curves from Cancer Trials Using External Information. *Med Decis Making.* 2017;37(4):353-366.

21. Sealey S, Edathodu A, Mukku SR. PHP109. International HTA Referencing – A Reality? *Value in Health.* 2014;17(3):A28.

22. Dickson R, Boland A, Duarte R, et al. EMA and NICE Appraisal Processes for Cancer Drugs: Current Status and Uncertainties. *Appl Health Econ Health Policy.* 2018;16(4):429-432.

23. National Institute for Health and Care Excellence (NICE). NICE Guidance: Technology appraisal guidance. https://www.nice.org.uk/guidance/published?type=ta. Accessed July 5, 2018.

24. National Institute for Health and Care Excellence (NICE). Technology appraisal guidance 405: Trifluridine–tipiracil for previously treated metastatic colorectal cancer. https://www.nice.org.uk/guidance/ta405. Published August 2016. Accessed July 5, 2018.

25. National Institute for Health and Care Excellence (NICE). Technology appraisal guidance 406: Crizotinib for untreated anaplastic lymphoma kinase-positive advanced non-small-cell lung cancer. https://www.nice.org.uk/guidance/ta406. Published September 2016. Accessed July 5, 2018.

26. National Institute for Health and Care Excellence (NICE). Technology appraisal guidance 411: Necitumumab for untreated advanced or metastatic squamous non-small-cell lung cancer. https://www.nice.org.uk/guidance/ta411. Published September 2016. Accessed July 5, 2018.

27. National Institute for Health and Care Excellence (NICE). Technology appraisal guidance 417: Nivolumab for previously treated advanced renal cell carcinoma. https://www.nice.org.uk/guidance/ta417. Published November 2016. Accessed July 5, 2018.

28. National Institute for Health and Care Excellence (NICE). Technology appraisal guidance 489: Vismodegib for treating basal cell carcinoma.

https://www.nice.org.uk/guidance/ta489. Published November 2017. Accessed July 5, 2018.

29. National Institute for Health and Care Excellence (NICE). Technology appraisal guidance 450: Blinatumomab for previously treated Philadelphia-chromosome-negative acute lymphoblastic leukaemia. https://www.nice.org.uk/guidance/ta450. Published June 2017. Accessed July 5, 2018.

30. National Institute for Health and Care Excellence (NICE). Technology appraisal guidance 462: Nivolumab for treating relapsed or refractory classical Hodgkin lymphoma. https://www.nice.org.uk/guidance/ta462. Published November 2017. Accessed July 5, 2018.

31. National Institute for Health and Care Excellence (NICE). Technology appraisal guidance 447: Pembrolizumab for untreated PD-L1-positive metastatic non-small-cell lung cancer. https://www.nice.org.uk/guidance/ta447. Published June 2017. Accessed July 5, 2018.

32. National Institute for Health and Care Excellence (NICE). Technology appraisal guidance 472: Obinutuzumab with bendamustine for treating follicular lymphoma refractory to rituximab. https://www.nice.org.uk/guidance/ta472. Published August 2017. Accessed July 5, 2018.

33. National Institute for Health and Care Excellence (NICE). Technology appraisal guidance 381: 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.

https://www.nice.org.uk/guidance/ta381. Published January 2016. Accessed Accessed July 4, 2018.

34. Ledermann J, Harter P, Gourley C, et al. Olaparib maintenance therapy in patients with platinum-sensitive relapsed serous ovarian cancer: a preplanned retrospective analysis of outcomes by BRCA status in a randomised phase 2 trial. *Lancet Oncol.* 2014;15(8):852-861.

35. Ledermann JA, Harter P, Gourley C, et al. Overall survival in patients with platinum-sensitive recurrent serous ovarian cancer receiving olaparib maintenance monotherapy: an updated analysis from a randomised, placebo-controlled, double-blind, phase 2 trial. *Lancet Oncol.* 2016;17(11):1579-1589.

36. European Medicines Agency (EMA). Lynparza: Product Information. http://www.ema.europa.eu/docs/en_GB/document_library/EPAR_-_Product_Information/human/003726/WC500180151.pdf. Accessed July 18, 2018.

37. Guyot P, Ades AE, Ouwens MJ, Welton NJ. Enhanced secondary analysis of survival data: reconstructing the data from published Kaplan-Meier survival curves. *BMC Med Res Methodol.* 2012;12:9.

38. Collett D. *Modelling survival data in medical research.* 2nd ed. Boca Raton, Fla. ; London: Chapman & Hall/CRC; 2003.

39. Royston P, Parmar MK. Flexible parametric proportional-hazards and proportional-odds models for censored survival data, with application to prognostic modelling and estimation of treatment effects. *Stat Med.* 2002;21(15):2175-2197.

40. Tappenden P, Harnan S, Ren S, et al. Olaparib for maintenance treatment of BRCA 1 or 2 mutated, relapsed, platinum-sensitive ovarian, fallopian tube and peritoneal cancer in people whose relapsed disease has responded to platinum-based chemotherapy: A Single Technology Appraisal. Produced by the School of Health and Related Research, University of Sheffield, commissioned by NIHR HTA Programme. https://www.nice.org.uk/guidance/ta381/documents/ovarian-fallopian-tube-and-peritoneal-cancer-brca-1-or-2-mutated-relapsed-platinumsensitive-olaparib-maintenance-id735-committee-papers12. Published March 2015. Accessed July 4, 2018.

41. Cancer Research UK. Ovarian cancer statistics. https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/ovarian-cancer/survival - heading-Three. Accessed Aug 22, 2018.

42. AstraZeneca UK Ltd. Olaparib as a monotherapy for the maintenance treatment of adult patients with platinum-sensitive relapsed (PSR) BRCA-mutated (germline and/or somatic) high-grade, serous epithelial ovarian, fallopian tube or primary peritoneal cancer who are in response (complete response or partial response) to platinum-based chemotherapy. Response to NICE Techinal Team & ERG Clarification Questions. https://www.nice.org.uk/guidance/ta381/documents/ovarian-fallopian-tube-and-peritoneal-cancer-brca-1-or-2-mutated-relapsed-platinumsensitive-olaparib-maintenance-id735-committee-papers12. Published February 2015. Accessed July 4, 2018.

43. Matulonis UA, Harter P, Gourley C, et al. Olaparib maintenance therapy in patients with platinum-sensitive, relapsed serous ovarian cancer and a BRCA mutation: Overall

survival adjusted for postprogression poly(adenosine diphosphate ribose) polymerase inhibitor therapy. *Cancer.* 2016;122(12):1844-1852.

44. Woods B, Sideris E, Palmer S, Latimer N, Soares M. NICE DSU Technical Support Document 19. Partitioned Survival Analysis for Decision Modelling in Health Care: A Critical Review. 2017. http://www.nicedsu.org.uk/. Published 2017. Accessed Aug 16, 2018.

45. AstraZeneca UK Ltd. Olaparib as a monotherapy for the maintenance treatment of adult patients with platinum-sensitive relapsed (PSR) BRCA-mutated (germline and/or somatic) high-grade, serous epithelial ovarian, fallopian tube or primary peritoneal cancer who are in response (complete response or partial response) to platinum-based chemotherapy. AstraZeneca submission to the National Institute for Health and Care Excellence. https://www.nice.org.uk/guidance/ta381/documents/ovarian-fallopian-tube-and-peritoneal-cancer-brca-1-or-2-mutated-relapsed-platinumsensitive-olaparib-maintenance-id735-committee-papers12. Published January 2015. Accessed July 10, 2018.

46. Strong M, Breeze P, Thomas C, Brennan A. SAVI-Sheffield Accelerated Value of Information. The University of Sheffield. http://savi.shef.ac.uk/SAVI/. Published 2018. Accessed August 22, 2018.

47. Strong M, Oakley JE, Brennan A. Estimating multiparameter partial expected value of perfect information from a probabilistic sensitivity analysis sample: a nonparametric regression approach. *Med Decis Making.* 2014;34(3):311-326.

48. Rohatgi A. WebPlotDigitizer. https://automeris.io/WebPlotDigitizer. Published 2018. Accessed July 5, 2018.

49. R Core Team. *R: A language and environment for statistical computing.* Vienna, Austria: R Foundation for Statistical Computing; 2013.

50. Jackson CH. flexsurv: A Platform for Parametric Survival Modeling in R. *J Stat Softw.* 2016;70.

51. NHS England Cancer Drugs Fund Team. Appraisal and funding of cancer drugs from July 2016 (including the new Cancer Drugs Fund): a new deal for patients, taxpayers and industry. https://www.england.nhs.uk/wp-content/uploads/2013/04/cdf-sop.pdf. Published July 2016. Accessed Dec 7, 2019.

52. National Institute for Health and Care Excellence (NICE). Guide to the processes of technology appraisal. https://www.nice.org.uk/Media/Default/About/what-we-do/NICE-guidance/NICE-technology-appraisals/technology-appraisal-processes-guide-apr-2018.pdf. Published April 2018. Accessed March 28, 2020.

53. Negrin MA, Nam J, Briggs AH. Bayesian Solutions for Handling Uncertainty in Survival Extrapolation. *Med Decis Making.* 2017;37(4):367-376.

54. Mahon R. *Temporal uncertainty in cost-effectiveness decision models*, University of York; 2014.

55. Jackson CH, Bojke L, Thompson SG, Claxton K, Sharples LD. A framework for addressing structural uncertainty in decision models. *Med Decis Making.* 2011;31(4):662-674.

56. Pujade-Lauraine E, Ledermann JA, Selle F, et al. Olaparib tablets as maintenance therapy in patients with platinum-sensitive, relapsed ovarian cancer and a BRCA1/2 mutation (SOLO2/ENGOT-Ov21): a double-blind, randomised, placebo-controlled, phase 3 trial. *Lancet Oncol.* 2017;18(9):1274-1284.

57. Bagust A, Beale SJ. Exploring the Effects of Early Censoring and Analysis of Clinical Trial Survival Data on Effectiveness and Cost-Effectiveness Estimation through a Case Study in Advanced Breast Cancer. *Med Decis Making.* 2018:272989X18790966.

58. Davies C, Briggs A, Lorgelly P, Garellick G, Malchau H. The "hazards" of extrapolating survival curves. *Med Decis Making.* 2013;33(3):369-380.

**TABLES**

**Table 1** Data availability and maturity for three main endpoints and three data-cuts, for 2L+ BRCAm population[35,36,45].

| **Time to event outcome** | **Nov 2012** | **Sept 2015** | **May 2016** |
|---|---|---|---|
| Median follow-up | 3 years | 6 years | 6.5 years |
| OS | V (52%) | V (70%) | V (73%) |
| TFST (a proxy for PFS) | V (74%) | V (82%) | X |
| TTD | V (87%) | X | X |

OS indicates overall survival; PFS, progression-free survival; TFST, time to first subsequent therapy or death; TTD, time to treatment discontinuation or death; V, data avilable; X, data not available.

Note: The availability was based on the availability of Kaplan-Meier curves which can be digitised to reconstruct IPD; figures in parentheses indicate the proportion of studied patients having events at the end of follow-up.

**Table 2** Cost-effectiveness analysis results

| | **Original data-cut** | **Latest data-cut** |
|---|---|---|
| ***Deterministic results*** | | |
| Incremental costs (£) | 36,706 | 36,972 |
| Incremental QALYs | 0.37 | 0.80 |
| ICER (per QALY) (£) | 99,570 | 46,339 |
| ***Probabilistic results*** | | |
| Incremental costs (£) (95% credible interval) | 36,588 (32,569, 40,525) | 36,848 (32,861, 40,588) |
| Incremental QALYs (95% credible interval) | 0.36 (-0.44, 1.14) | 0.80 (0.05, 1.63) |
| ICER (per QALY) (£) | 101,467 | 45,787 |
| Incremental NMB* | -25,770 | -12,705 |
| (95% credible interval) | (-49,204, -2,224) | (-35,444, 11,386) |
| ***Overall EVPI**** | | |
| Individual EVPI (per person) (£) | 68 | 978 |
| Population EVPI (over five years) (£) | 152,141 | 2,201,146 |

ICER indicates incremental cost-effectiveness ratio; QALY, quality-adjusted life year; NMB, net monetary benefit.

* EVPI and NMB were estimated at the threshold of £30,000 per QALY gained

**Table 3** Scenario analysis

| **ICER (per QALY) (£)** | **Original data-cut** | **Latest data-cut** |
|---|---|---|
| **Deterministic base-case** | **99,570** | **46,339** |
| ***OS (TFST and TTD: best-fitting models*)*** | | |
| Exponential | 61,684 | 39,137 |
| Weibull | 127,006 | 41,837 |
| Gompertz | 148,952 | 40,320 |
| Log-normal | - | 40,410 |
| Log-logistic | 110,194 | 43,195 |
| Generalised gamma | 280,487 | 36,535 |
| Default spline (normal, m=3) | - | 33,004 |

ICER indicates incremental cost-effectiveness ratio; OS, overall survival; QALY, quality-adjusted life year; TFST, time to first subsequent therapy or death; TTD, time to treatment discontinuation or death.

*The best-fitting models for the original data-cut were log-normal, log-normal and log-logistic for OS, TFST and TTD respectively; best-fitting models for the latest data-cut were user-specified spline, user-specified spline and log-logistic for OS, TFST, TTD respectively.

## FIGURES

**Figure 1** Proportion of death in STAs. (A) Mature STAs. (B) Immature STAs. NR indicates that the proportion of death was not reported; STA, single technology appraisal.

* The proportion/number of death was only reported for primary data-cut in this TA; however, the company has submitted data with newer data-cuts for which the company did not report proportion/number of death. Thus, here we summarized based on primary data-cut for this specific TA.

**Figure 2** Comparison of predicted and actual long-term OS curves. Actual indicates actual long-term Kaplan-Meier curve; Projected, predicted curves using short-term data.