
Article:
Bullement, A., Latimer, N.R. and Bell Gorrod, H. (2019) Survival extrapolation in cancer immunotherapy: a validation-based case study. Value in Health, 22 (3). pp. 276-283. ISSN: 1098-3015

https://doi.org/10.1016/j.jval.2018.10.007

Survival extrapolation in cancer immunotherapy: a validation-based case study

Bullement A, Latimer NR, Bell Gorrod H.

Abstract

243/250 words

Objectives: Immune-checkpoint inhibitors may provide long-term survival benefits via a cured proportion, yet data are usually insufficient to prove this upon submission to health technology assessment bodies. We revisit the National Institute for Health and Care Excellence assessment of ipilimumab in melanoma (TA319). We use updated data from the pivotal trial to assess the accuracy of the extrapolation methods used and compare these to previously-unused techniques to establish whether an alternative extrapolation may have provided more accurate survival projections.

Methods: We compare projections from the piecewise survival model used in TA319 and those produced by alternative models (fit to trial data with minimum follow-up of 3 years), to a longer-term data-cut (5-year follow-up). We also compare projections to external data to help assess validity. Alternative approaches considered were parametric, spline-based, mixture, and mixture-cure-models.

Results: Only the survival model used in TA319 and a mixture-cure model provided 5-year survival predictions close to those observed in the 5-year follow-up dataset. Standard parametric, spline, and non-curative-mixture models substantially under-estimated 5-year survival. Survival estimates from the TA319 model and the mixture-cure model diverge considerably after 5 years and remain unvalidated.

Conclusions: In our case study, only models that incorporated an element of external information (through a cure fraction combined with background mortality rates or using registry data) provided accurate estimates of 5-year survival. Flexible models that were able to capture the complex hazard functions observed during the trial, but which did not incorporate external information, extrapolated poorly.

Highlights

- Extrapolations of survival for cancer immunotherapies are associated with substantial uncertainty, and often emerge as a key concern when cancer immunotherapies are appraised by the National Institute for Health and Care Excellence (NICE).
- No previous study has attempted to validate extrapolations from early data-cuts of cancer immunotherapy trials via comparison to later data-cuts.
- Through comparing projections from an early data-cut to actual survival from a later data-cut, we show that in this case study seemingly optimistic extrapolations may in fact be credible. Assessing extrapolation fit to new data is useful for providing evidence on the performance of extrapolation methods.

Main body

3,984/4,000 words

Introduction

Immunotherapy is an area of ever-evolving research in the treatment of cancer, most recently with the development of immune-checkpoint inhibitors (ICIs) – a type of monoclonal antibody which aims to enhance the immune system to destroy cancer cells. (1-3) ICIs may result in a proportion of patients achieving long-term survival outcomes, sometimes referred to as a "statistically-cured" fraction. (4-11) However, ICI registrations are characterised by

clinical trials with limited follow-up, and hence there is a great deal of uncertainty as to whether there truly is a "statistically-cured" fraction yet to be observed. Decision-makers are consequently placed in a difficult position, given that the estimation of survival (and whether or not this includes a "statistically-cured" fraction) plays a huge role in determining an intervention's clinical and cost effectiveness.

The National Institute for Health and Care Excellence (NICE) published guidance on the first licensed ICI ipilimumab for patients with treatment-naïve advanced melanoma in July 2014 (TA319). (12-14) As part of TA319, the manufacturer submitted a cost-effectiveness analysis (CEA) based on data from the pivotal CA184-024 trial with a minimum follow-up period of 37 months (henceforth referred to as 3 years). (15) The survival extrapolation used in TA319 combined registry data for patients not treated with ipilimumab with CA184-024 data in a piecewise approach. Following publication of TA319, further follow-up data (5-year minimum follow up) have been published from CA184-024. (16, 17)

The NICE Decision Support Unit (DSU) first published guidance regarding survival analysis and extrapolation using patient-level data (PLD) in June 2011 in the form of Technical Support Document (TSD) 14. (18) TSD-14 focuses mainly on six "standard" parametric distributions (exponential, Weibull, Gompertz, log-normal, log-logistic and generalized gamma). Other methods are also discussed in TSD-14 (e.g. piecewise and flexible parametric models) though only in brief. More recently, mixture- and spline-based models have been presented on several occasions in the literature. (19-21) These newer methods were not well known in the world of health technology assessment (HTA) around the time of TA319.

The aim of this study is to assess the survival predictions obtained from the piecewise model used in TA319 compared to those from a range of alternative models. Models are fit to the 3-year data-cut (available at the time of TA319), and predicted survival is compared to

observed survival in the 5-year data-cut. This allows an assessment of how well each model estimates 5-year survival. Projections are also compared to long-term survival trends in melanoma exhibited in registry data to provide information on their credibility beyond 5 years.

Methods

The dataset

The CA184-024 trial compared ipilimumab + dacarbazine (IPI+DTIC) to placebo + dacarbazine (PBO+DTIC) in treatment-naïve advanced melanoma. (15) CA184-024 was a multi-centre, randomised, double-blind, Phase III study of 502 adult stage III/IV melanoma patients. (15) PLD for overall survival were recreated from published 3-year and 5-year Kaplan-Meier (KM) curves using an algorithm by Guyot et al. (22) The recreated PLD were required in order to fit and subsequently compare alternative survival extrapolations, as the "true" PLD are not publicly available. In the 3-year data-cut 47 patients treated with IPI+DTIC and 28 patients treated with PBO+DTIC remained at risk at 3 years. In the 5-year data-cut 40 patients treated with IPI+DTIC and 18 patients treated with PBO+DTIC were at risk at 5 years.

Candidate models

The model used in TA319 incorporated three piecewise components: (1) the KM curve from CA184-024 to inform survival until 2 years; (2) a log-normal curve fitted to patients alive in CA184-024 beyond 2 years to inform survival from 2 until 5 years; and (3) a Weibull curve fitted to long-term registry data from the American Joint Committee on Cancer (AJCC) to inform survival beyond 5 years. (12) The same approach was adopted for both treatment arms, such that the hazard of death for all patients was equal after 5 years.

In addition to the TA319 piecewise model, we considered "standard" parametric survival models, Royston and Parmar spline-based models, and mixture models. Standard parametric survival and piecewise models are well documented in published literature, and therefore are not discussed in detail here. (18, 23-26) However, the other candidate models are less well known, and therefore brief overviews of the models are provided:

Spline-based models consider extensions of the Weibull, log-normal and log-logistic distributions where natural cubic splines (piecewise polynomials) are used to smooth between piecewise sections of the model. (27) Gibson et al. (20) have previously considered these models in the context of ICIs – fitting standard parametric and spline-based models to progression-free survival data from the CheckMate-067 trial of nivolumab and ipilimumab in advanced melanoma and concluding that the spline-based models were able to characterise a flattening of the survival curve (a "plateau") observed within the data.

Mixture models have been considered within the literature to reflect the apparent heterogeneity evident in populations of patients, whereby separate groups have different underlying survival distributions. Chen (28) and Othus et al. (21) present parametric mixture-cure models (MCMs) as a method to quantify the expectation that a group of ICI patients may achieve long-term survival. The parametric MCMs consider two separate survival trajectories, estimating a "statistically-cured" proportion of patients (determined via logistic regression). That is, a proportion of patients die from their disease, and a proportion of patients will follow the survival trajectory that would be expected for the age-adjusted general population and do not die of their disease. We considered log-normal and Weibull MCM's, and incorporated background mortality from United Kingdom (UK) age- and sex-matched life tables from the Office for National Statistics (ONS), assuming baseline mean patient characteristics from CA184-024. (29)

Another example of a mixture model is the mixture-Weibull model (MWM). Sánchez et al. (30) utilised the MWM for the survival of immunotherapy-treated advanced non-small-cell lung cancer patients, and Crowther et al., (2016) (31) used the MWM to fit to "complex survival data", with an example in breast cancer. The MWM demonstrates improved visual fit versus the standard Weibull model in the latter of these examples. These mixture models permit different survival distributions to exist within the data, but do not assume a "statistically-cured" fraction. We considered the MWM in this study but no other non-cure-based mixture models.

Selecting potentially appropriate models

In order to assess the potential suitability of the models, we assessed the recreated PLD using a variety of techniques. First, we evaluated the underlying hazard function in order to establish the suitability of models that assume proportional hazards (PH models) or constant acceleration factors (accelerated failure time [AFT] models). Second, we used the hazard function to assess the applicability of standard parametric forms (such as the exponential, which assumes a constant hazard rate). Third, we used the hazard function to identify turning points to inform the appropriate number of knots in the spline-based models. Models considered potentially suitable were fitted to the 3-year re-created PLD using the statistical software packages R and Stata. (32, 33)

Assessing model predictions

Following model fitting, an assessment of each model was undertaken. The fit to the observed data (via visual inspection and/or residual analysis) was first considered. Next, the plausibility and credibility of survival extrapolation was ascertained through comparing trajectories to external data and background mortality rates. Finally, the statistical goodness-of-fit (using Akaike or Bayesian Information Criteria [AIC/BIC]) was used to determine

models that incorporate an inefficient use of additional parameters to inform their estimation of survival. This process of selecting appropriate survival models is in line with TSD-14 recommendations. (18)

In TA319 long-term survival data for advanced melanoma patients treated with non-ICI treatment available from the AJCC were used to inform survival projections. (34) These data provided estimates of 15-year survival which may be considered useful when ascertaining whether or not survival projections demonstrate clinical plausibility. Baseline characteristics for patients in CA184-024 are published however the corresponding data for AJCC patients are unreported. (15, 34) Therefore, the comparability of the AJCC and CA184-024 data is unclear.

To assess the survival predictions made by the different fitted models we compared these to the 5-year CA184-024 data-cut, the AJCC data, and mortality rates for the UK age- and sex-matched general population. (29) In particular, the AJCC data were used to assess the plausibility of survival extrapolations for the control arm; noting the previous reporting limitations and acknowledging the ages of the respective studies.

Results

Survival observed in CA184-024

The re-created KM curves are presented in Figure 1 for the 3- and 5-year data-cuts. Median survival for patients treated with PBO+DTIC in the 3-year data-cut was approximately 9.1 versus 11.2 months for IPI+DTIC. (15) 5-year survival in the 5-year data-cut was 18.2% for IPI+DTIC patients versus 8.8% for PBO+DTIC patients. (16)

Estimated smoothed hazard functions for both data-cuts are also presented in Figure 1. Hazards originally increased, before decreasing in the longer-term in both treatment arms.

There was an apparent increase close to the end of trial-follow-up in the 3-year data-cut for patients treated with IPI+DTIC, but this is highly uncertain due to the small number of patients at risk at this point (and indeed this increase was not present in the 5-year data-cut).

In CA184-024, patients were required to have a life expectancy of at least 16 weeks prior to treatment initiation, (15) and so a low initial hazard of death may be expected. However, given the severity of advanced melanoma, the hazard of death would be expected to increase following the start of the study. In the longer term however, patients with a better prognosis (i.e. responders) would still be alive, and so the hazard of death for the surviving group of patients would fall. Further in the future, the hazard of death may increase again as patients age (though this may be difficult to determine due to trial follow-up and patient numbers).

Based upon the hazard functions estimated in the 3-year data-cut for both treatment arms we deemed that only parametric models that can represent increasing and then decreasing (i.e. non-monotonic) hazards were appropriate. Models that assume either constant or unidirectional (monotonic) hazards would provide a poor representation of the observed data. This meant the exponential, Gompertz and Weibull distributions were inappropriate for consideration.

The smoothed hazard function also demonstrates that the treatment effect does not appear to be proportional (this was also suggested by log-cumulative hazard and quantile-quantile plots, not presented here) and therefore models were fitted to treatment arms independently. Based on CA184-024 data, and the assessment of the corresponding smoothed hazard plot, the following candidate models were considered potentially appropriate and were fitted: (1) Log-normal, log-logistic, generalised gamma and generalised-F parametric survival models; (2) Spline-based models; (3) MCMs; and (4) MWM.

The number of turning points in the smoothed hazard plot was used to determine the number of knots for the spline-based models. Models with one or two knots were considered as a maximum of two turning points were evident in the hazard plot. 3-knot models were considered, though provided substantially poorer statistical goodness-of-fit scores and are therefore not presented. The best fitting one- and two-knot spline-based models (in terms of AIC and BIC) were the odds- and hazard-based splines, respectively; and so were considered henceforth. The knot locations were determined according to percentiles of the uncensored survival times, in line with recommendations. (27)

The statistical goodness-of-fit for each survival model is presented in Table 1. All models were also considered in terms of their visual fit to the observed data, shown in Figure 2. Each of the models exhibited a reasonable fit to the observed data, though by year 5 there is a clear spread of projected survival estimates ranging from 10%-18%.

Survival beyond the observed period of CA184-024

Predicted survival proportions at 3, 5, 10, 15 years, and mean survival estimates associated with each model/ data source are presented in Table 2. The corresponding survival curves are presented in Figure 2, alongside KM curves from the 5-year data-cut of CA184-024 and AJCC data.

The TA319 piecewise model yielded a mean survival estimate of 3.59 life-years for IPI+DTIC patients and 2.50 life-years for PBO+DTIC patients – greater than all the other models except for the two MCMs.

Only the TA319 piecewise model and the MCM's provide 5-year survival probabilities close to those observed for IPI+DTIC in the 5-year data-cut (17.5%, 17.4% and 14.4% for the TA319 model, Weibull MCM and log-normal MCM respectively, compared to 18.1%

observed in the 5-year data-cut). The parametric models, MWM and spline-based models under-predict 5-year survival for IPI+DTIC (ranging from 10.1% to 13.2%). Beyond 5 years the survival predicted for IPI+DTIC by the TA319 model and the MCMs diverge considerably. The flattened shape of the survival curves predicted by the MCMs appears to be more in line with survival at 6 years observed in the 5-year data-cut from CA184-24, but beyond this point the IPI+DTIC curves remain unvalidated.

In contrast, the TA319 model and the MCMs appear to over-estimate 5-year survival for the PBO+DTIC group (11.0%, 11.1% and 9.1% for the TA319 model, Weibull MCM and log-normal MCM respectively, compared to 8.1% observed in the 5-year data-cut). The parametric and spline-based models provide under-estimates of 5-year survival for PBO+DTIC, whereas the MWM provides an accurate estimate. However, the AJCC data suggests that the survival curve may be expected to flatten somewhat even for the PBO+DTIC group – the MWM does not follow this trend. The log-normal MCM and the TA319 model are the only models that appear to project survival in a way that is consistent with the AJCC data.

Figure 3 (see Supplementary Material) presents the estimated hazard of death for both groups of patients projected by each model, as well as those estimated based on the observed data from CA184-024 (5-year data-cut) and ONS background mortality rates. All models except the MWM, the MCMs and the TA319 piecewise model demonstrate decreasing hazards beyond 5 years. The MWM and the TA319 models both include a Weibull component which exhibits monotonically increasing hazards. For the MCMs the hazard increases in the longer-term because increasing age-related hazards are accounted for. Only the MCMs result in extrapolations in which mortality beyond approximately 5-10 years is represented by

background mortality rates – all other approaches exhibit hazards that are substantially higher than background mortality.

Discussion

The MCMs and the piecewise approach used in TA319 were the only models that provided extrapolations for IPI+DTIC that were aligned with the 5-year data-cut from CA184-024. These also predicted the highest mean survival for IPI+DTIC. Flexible models, such as the spline-based models and the MWM, were able to represent non-monotonic hazards but appeared to over-estimate longer-term hazards.

The AJCC data suggest 15-year survival for patients not treated with IPI+DTIC is approximately 4.2%. (34) This is of limited use for assessing survival predictions for patients treated with IPI+DTIC, but may be considered to represent a lower bound for potentially credible extrapolations. Only the two MCMs and the TA319 piecewise model predicted 15-year survival of at least 4.2% for both treatment arms.

For PBO+DTIC patients, several methods predicted 5-year survival in line with the 5-year data-cut. However, predictions of survival differed substantially in the longer-term, with 15-year survival estimates ranging from 0.5% (MWM) to 9.8% (Weibull MCM). Several models exhibited a distinct lack of face validity, predicting 10-year survival of less than 3% and all models except the MCMs and the TA319 piecewise model predicted lower long-term survival than that observed in the AJCC data.

While CA184-024 data show survival for PBO+DTIC to be lower than the AJCC survival curve in the short-term, the gap appeared to narrow according to the 5-year data-cut. It appears that the Weibull MCM may over-estimate longer-term survival for PBO+DTIC patients, whereas the TA319 piecewise approach and the log-normal MCM appear to provide

more credible estimates. However, the aforementioned limitations of the AJCC data remain. As such, the estimate of 15-year survival for PBO+DTIC patients of 10% provided by the Weibull MCM may not be completely unrealistic.

In this study we identified that the underlying hazard functions estimated for the 3-year data-cut were complex, and that only models that could account for turning points in the hazard function would yield appropriate extrapolations. Accordingly, we fit a range of potentially appropriate models, but found that several of these did not provide accurate estimates of 5-year survival based upon a subsequent data-cut. Only models that incorporated external information (MCMs through background mortality data, and the TA319 piecewise model through the use of mortality rates from the AJCC registry) were able to accurately predict 5-year survival whilst also providing a survival curve with the shape expected based upon long-term registry data. This strengthens previous suggestions around the use of external information in survival extrapolation. (35, 36)

Our analysis also provides further evidence illustrating the substantial impact that the extrapolation method can have on mean survival estimates, which in turn are likely to have an important effect on CEA results. Within the context of a treatment with potential long-term survival benefits, the importance of selecting an appropriate survival extrapolation method is heightened.

Our case study appears to offer support for the use of MCM's, which show one of the best fits to the 5-year data-cut, and exhibit estimates of longer-term hazards in line with registry data. However, these models also provided the largest estimates of survival and at the time HTA decisions are made there may be an understandable reluctance to believe such optimistic estimates. In addition, we have demonstrated that the parameterisation of an MCM has an important impact on survival projections – the Weibull and log normal MCMs

produced substantially different extrapolations – illustrating how sensitive extrapolations are to the "statistically-cured" fraction estimated, which may also vary by parameterisation. Given the small patient numbers that usually remain at the end of follow-up in clinical trials – which is key for estimating the "statistically-cured" fraction – caution is certainly required when using MCMs in a clinical trial context.

In TA319 a piecewise modelling approach was taken which appears to have produced accurate estimates of 5-year survival, and results in longer estimates of mean survival than most other models. However, our analyses suggest that even these estimates may have been pessimistic for IPI+DTIC, and optimistic for PBO+DTIC patients. The specification of a piecewise model requires the analyst to specify relevant cut-points at which to fit different models (in this case study, the model was cut at 2 and 5 years). Davies et al. (26) demonstrated how the choice of these cut-points can have a profound effect on the overall extrapolation produced, and the choice of cut-points for piecewise models is frequently criticised within NICE technology appraisals (including TA319). (12, 37, 38) The same criticism may be raised for spline-based models (in relation to knot locations) – whilst this may not be critical for model fit, (27) it may have an important effect on extrapolations.

Despite their limitations, we have shown that the piecewise approach and MCMs appear to have performed relatively well in our case study. We contend that it is not a coincidence that these were the only two methods that incorporated external information in some form. Other approaches may have provided similarly plausible longer-term survival estimates if they were adapted to include similar external information. For instance, spline-based models can be fitted in a relative survival framework, whereby "disease-specific mortality" and "other-cause mortality" are modelled separately, with information on background mortality incorporated for the latter component. (39, 40) Where "disease-specific mortality" rates fall to zero a

statistical cure is essentially modelled, potentially resulting in the kind of survival curve plateau observed in the 5-year data-cut and in the AJCC registry data. A limitation of our work is that we did not attempt to fit such models – but further use of these in an HTA and extrapolation context should be explored.

Recreated PLD were used to fit survival models in this study, and so extrapolation methods that require "true" PLD were not considered (for example, response-based models discussed by Huang *et al.*) (19) Access to "true" PLD would allow for improved accuracy of model fitting, adjustment for covariates and the ability to consider more complex models. However, the use of recreated PLD does not impede the fitting of survival models in general, and consequently provides a sufficient basis to compare survival projections.

Our study demonstrates that it is useful to collect longer-term information from clinical trials in order that the accuracy of previous extrapolations can be assessed – such updates should be encouraged whenever long-term survival is expected but not proven. (6) Further data collection from the same trial may be requested by NICE relating to a conditional approval via the Cancer Drugs Fund, or a periodic review of guidance (for TA319, this was scheduled 2 years following publication of the Final Appraisal Determination). (12)

However, whilst continued publication of updated survival information is highly commendable, the 5-year data-cut from CA184-024 unavoidably remains limited to a relatively short time period, compared to the potential survival times of "statistically-cured" survivors. It is striking that even models that produce similar estimates of survival at 5 years go on to project considerably different mean survival, for both treatment arms. In particular, even in the 5-year data-cut the existence of a "statistically-cured" fraction is still unclear, though published literature suggest that conditional survival appears to improve over time in melanoma – Xing et al. (41) reported that conditional 5-year survival for stage IV melanoma

patients increased from 19% upon diagnosis to 84% if patients survive until 5 years, based on data from the Surveillance, Epidemiology, and End Results (SEER) Program (though these patients did not receive ICI treatment).

It should be further acknowledged that this study reports the findings of one case study in advanced melanoma. We have shown that cure-based models may represent a useful tool for projecting survival when long-term survival is expected in a proportion of patients. However, the presence of a "cure fraction" is not evidence of a "true" cure. Publication of updated data-cuts from clinical trials in which long-term survival is expected would allow further investigation of whether or not a true cure is evident, which in turn may provide additional justification for undertaking survival modelling using cure-based models. Subsequently, decision-makers may have more confidence in basing decisions on analyses that assume that a proportion of patients are truly cured.

TSD-14 was last updated in 2013. (18) Since then, a number of ICIs have been assessed by NICE which have adopted increasingly complex extrapolation methods – recently NICE TA590 of nivolumab for unresectable/ metastatic urothelial cancer included the use of a response-based model (combined with a landmark analysis to overcome immortal time bias) as standard models would "fail to capture the changes in hazard over time". (42) Our case study provides an overview of different extrapolation methods (some of which are not discussed within TSD-14), which may aid the future update of guidance. However, it was not possible to consider all methods and only data from CA184-024 were considered, hence further research is required to provide comprehensive guidance for the fitting of complex survival models.

Conclusions

Our study demonstrates how updated clinical trial data-cuts can be used to assess the accuracy of extrapolations made by fitting survival models to earlier data-cuts. This provides useful information on which models appear to extrapolate well, and which appear to extrapolate inaccurately. We only consider one case study, but the findings are useful more broadly, and are of particular relevance in diseases where long-term survival is expected.

Flexible models that are able to accurately represent complex hazard functions are likely to provide a good fit to observed data but may not extrapolate accurately – in our study these methods under-estimated 5-year survival for IPI+DTIC. Models that incorporated external information – either by assuming a "statistically-cured" fraction and incorporating background mortality rates, or by extrapolating in line with longer-term registry data – provided more accurate estimates of 5-year survival, demonstrating the importance of including relevant external information when available. However, even these models projected beyond 5 years very differently, demonstrating the need to continue to collect survival data in order to further aid understanding of long-term survival.

References

1. Dine J, Gordon R, Shames Y, et al. Immune Checkpoint Inhibitors: An Innovation in Immunotherapy for the Treatment and Management of Patients with Cancer. Asia Pac J Oncol Nurs. 2017; 4: 127-35.

2. Cancer Research UK (CRUK). What are 'checkpoint' immunotherapy drugs? (YouTube video). 2015.

3. Couzin-Frankel J. Cancer Immunotherapy. Science. 2013; 342: 1432.

4. Lynch TJ, Bondarenko I, Luft A, et al. Ipilimumab in combination with paclitaxel and carboplatin as first-line treatment in stage IIIB/IV non-small-cell lung cancer: results from a randomized, double-blind, multicenter phase II study. J. Clin. Oncol. 2012; 30: 2046-54.

5. Hodi FS, O'Day SJ, McDermott DF, et al. Improved Survival with Ipilimumab in Patients with Metastatic Melanoma. N. Engl. J. Med.. 2010; 363: 711-23.

6. Schadendorf D, Hodi FS, Robert C, et al. Pooled Analysis of Long-Term Survival Data From Phase II and Phase III Trials of Ipilimumab in Unresectable or Metastatic Melanoma. J. Clin. Oncol. 2015; 33: 1889-94.

7. Reck M, Rodríguez-Abreu D, Robinson AG, et al. Pembrolizumab versus Chemotherapy for PD-L1–Positive Non–Small-Cell Lung Cancer. N. Engl. J. Med. 2016; 375: 1823-33.

8. Brahmer J, Reckamp KL, Baas P, et al. Nivolumab versus Docetaxel in Advanced Squamous-Cell Non–Small-Cell Lung Cancer. N. Engl. J. Med. 2015; 373: 123-35.

9. Kaufman HL, Russell J, Hamid O, et al. Avelumab in patients with chemotherapy-refractory metastatic Merkel cell carcinoma: a multicentre, single-group, open-label, phase 2 trial. Lancet Oncol. 2016; 17: 1374-85.

10. Rosenberg JE, Hoffman-Censits J, Powles T, et al. Atezolizumab in patients with locally advanced and metastatic urothelial carcinoma who have progressed following treatment with platinum-based chemotherapy: a single-arm, multicentre, phase 2 trial. Lancet. 2016; 387: 1909-20.

11. Antonia SJ, Villegas A, Daniel D, et al. Durvalumab after Chemoradiotherapy in Stage III Non–Small-Cell Lung Cancer. N. Engl. J. Med. 2017; 377: 1919-29.

12. National Institute for Health and Care Excellence (NICE). TA319: Ipilimumab for previously untreated unresectable malignant melanoma. 2014.

13. Leach DR, Krummel MF, Allison JP. Enhancement of Antitumor Immunity by CTLA-4 Blockade. Science. 1996; 271: 1734-36.

14. European Medicines Agency (EMA). EPAR summary for the public: Yervoy (ipilimumab). 2013.

15. Robert C, Thomas L, Bondarenko I, et al. Ipilimumab plus Dacarbazine for Previously Untreated Metastatic Melanoma. N. Engl. J. Med. 2011; 364: 2517-26.

16. Maio M, Grob J-J, Aamdal S, et al. Five-Year Survival Rates for Treatment-Naive Patients With Advanced Melanoma Who Received Ipilimumab Plus Dacarbazine in a Phase III Trial. J. Clin. Oncol. 2015; 33: 1191-96.

17. Maio M, Bondarenko I, Robert C, et al. Four-year Survival Update for Metastatic Melanoma (MM) Patients (pts) (IPI) + Dacarbazine (DTIC) Treated With Ipilimumab in Phase 3 Study CA184-024. 37th European Society for Medical Oncology (ESMO) Congress. Vienna: Austria, 2012.

18. Latimer N. NICE DSU Technical Support Document 14: Undertaking survival analysis for economic evaluations alongside clinical trials - extrapolation with patient-level data. 2011.

19. Huang M, Latimer NR, Zhang Y, et al. Estimating the Long-Term Outcomes Associated With ImmunoOncology Therapies: Challenges and Approaches for Overall Survival Extrapolations. Value & Outcomes Spotlight. 2018; 4.

20. Gibson E, Koblbauer I, Begum N, et al. Modelling the Survival Outcomes of Immuno-Oncology Drugs in Economic Evaluations: A Systematic Approach to Data Analysis and Extrapolation. PharmacoEconomics. 2017.

21. Othus M, Bansal A, Koepl L, et al. Accounting for Cured Patients in Cost-Effectiveness Analysis. Value Health. 2017; 20: 705-09.

22. Guyot P, Ades A, Ouwens M, et al. Enhanced secondary analysis of survival data: reconstructing the data from published Kaplan-Meier survival curves. BMC Med. Res. Methodol. 2012; 12: 9.

23. Tremblay G, Livings C, Crowe L, et al. Determination of the most appropriate method for extrapolating overall survival data from a placebo-controlled clinical trial of lenvatinib for progressive, radioiodine-refractory differentiated thyroid cancer. Clin. Outcomes Res. 2016; 8: 323-33.

24. Collett D. Modelling survival data in medical research. 2 ed.: Boca Raton: Chapman & Hall, 2003.

25. Latimer NR. Survival Analysis for Economic Evaluations Alongside Clinical Trials—Extrapolation with Patient-Level Data: Inconsistencies, Limitations, and a Practical Guide. Med. Decis. Making. 2013; 33: 743-54.

26. Davies A, Briggs A, Schneider J, et al. The Ends Justify the Mean: Outcome Measures for Estimating the Value of New Cancer Therapies. Health Outcomes Res. Med. 2012; 3: e25-e36.

27. 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: 2175-97.

28. Chen T-T. Predicting analysis times in randomized clinical trials with cancer immunotherapy. BMC Med. Res. Methodol. 2016; 16: 12.

29. Office for National Statistics (ONS). Dataset: National Life Tables: United Kingdom. 2016.

30. Sánchez L, Muchene L, Luaces P, et al. Evaluating the effect of immunotherapy in advanced non-small-cell lung cancer patients using two components mixture model. ISPOR 17th Annual European Congress. Amsterdam: The Netherlands, 2014.

31. Crowther MJ, Lambert PC. Simulating complex survival data. Stata Nordic and Baltic Users' Group Meeting. Stockholm: Sweden, 2011.

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

33. StataCorp. Stata Statistical Software: Release 13. College Station, TX: StataCorp LP. 2013.

34. Balch CM, Buzaid AC, Soong SJ, et al. Final version of the American Joint Committee on Cancer staging system for cutaneous melanoma. J. Clin. Oncol. 2001; 19: 3635-48.

35. Guyot P, Ades AE, Beasley M, et al. Extrapolation of Survival Curves from Cancer Trials Using External Information. Med. Decis. Making. 2017; 37: 353-66.

36. 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: 377-90.

37. National Institute for Health and Care Excellence (NICE). Technology appraisal guidance ID925: Alectinib for untreated anaplastic lymphoma kinase-positive advanced non-small-cell lung cancer. 2018.

38. Andersson TM, Dickman PW, Eloranta S, et al. Estimating the loss in expectation of life due to cancer using flexible parametric survival models. Stat. Med. 2013; 32: 5286-300.

39. Andersson TM, Dickman PW, Eloranta S, et al. Estimating and modelling cure in population-based cancer studies within the framework of flexible parametric survival models. BMC Med. Res. Methodol. 2011; 11: 96.

40. Dickman PW, Sloggett A, Hills M, et al. Regression models for relative survival. Stat. Med. 2004; 23: 51-64.

41. Xing Y, Chang GJ, Hu CY, et al. Conditional survival estimates improve over time for patients with advanced melanoma: results from a population-based analysis. Cancer. 2010; 116: 2234-41.

42. National Institute for Health and Care Excellence (NICE). TA590: Nivolumab for treating metastatic or unresectable urothelial cancer after platinum-based chemotherapy. 2018.

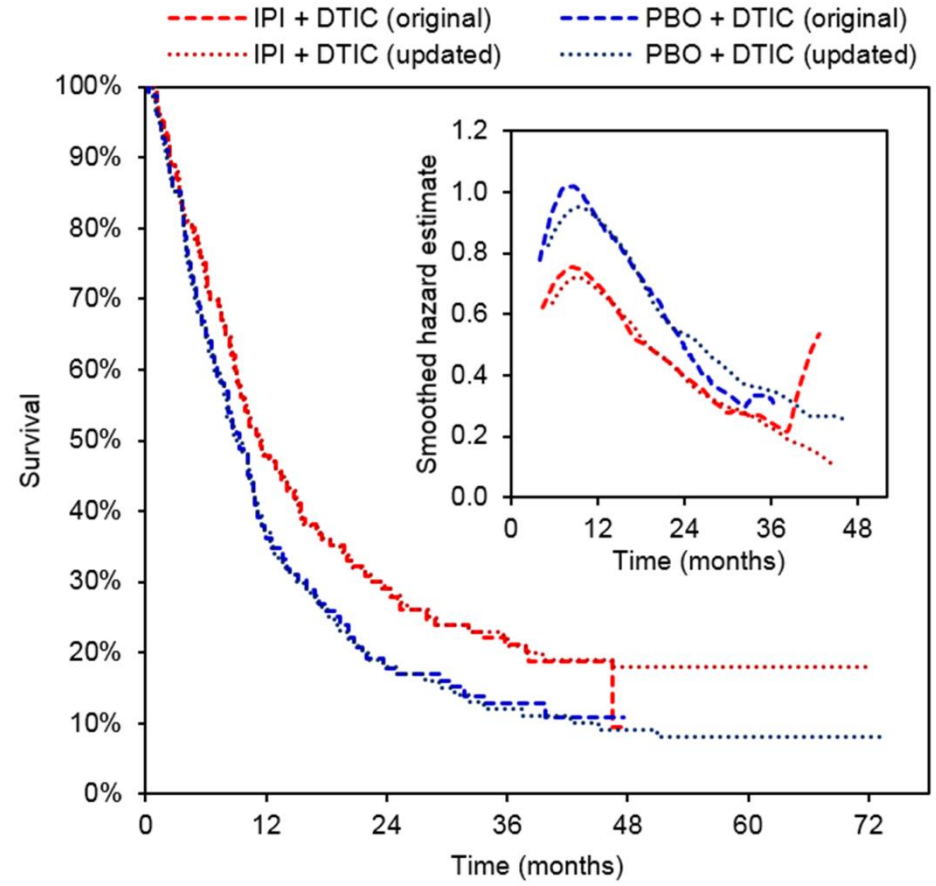
Table 1 – Statistical goodness of fit for survival models

Parameterisation	AIC	BIC
(1) Parametric survival models		
Generalised-F	3,105.29	3,133.50
Generalised Gamma	3,108.95	3,130.10
Log-logistic	3,108.99	3,126.62
Log-normal	3,107.53	3,121.63
(2) Spline-based models		
Odds – 1-knot	3,100.05	3,116.93
Normal – 1-knot	3,106.56	3,123.43
Hazard – 1-knot	3,101.06	3,117.94
Odds – 2-knots	3,100.45	3,121.54
Normal – 2-knots	3,099.97	3,121.06
Hazard – 2-knots	3,098.99	3,120.09
(3) Mixture-cure models		
Weibull	3,053.43	3,063.36
Log-normal	3,024.60	3,034.70
(4) Mixture-Weibull model		
Mixture-Weibull model	3,110.93	3,113.03

Table 2 – Mean survival estimates for each data source and survival extrapolation

Data source		Survival (%)				RMST (years)	Maximum time (years)
		3 years	5 years	10 years	15 years		
CA184-024 (Original)	IPI + DTIC PBO + DTIC	21.2 12.9	-	-	-	1.4747 1.2106	4
CA184-024 (Updated)	IPI + DTIC PBO + DTIC	22.1 12.0	18.1 8.1	-	-	1.7957 1.3591	6
AJCC registry		16.1	10.6	7.0	4.2	1.9838	15
Extrapolation		**Survival (%)**				**Mean survival (years)**	
		3 years	**5 years**	**10 years**	**15 years**	**Total**	**Δ**
TA319 PM	IPI + DTIC PBO + DTIC	22.2 14.9	17.5 11.0	11.1 6.9	7.0 4.4	3.5597 2.4995	1.0602
Weibull MCM	IPI + DTIC PBO + DTIC	20.4 12.3	17.4 11.1	16.4 10.5	15.1 9.8	5.4738 3.7267	1.7471
Log-normal MCM	IPI + DTIC PBO + DTIC	20.8 13.5	14.4 9.1	10.5 6.8	9.2 6.1	4.0389 2.8750	1.1639
Generalised gamma	IPI + DTIC PBO + DTIC	20.8 12.5	12.3 5.8	5.4 1.6	3.1 0.6	2.6843 1.5717	1.1127
Spline - 2 knot - Hazard	IPI + DTIC PBO + DTIC	21.0 13.7	13.2 7.9	5.6 2.9	2.9 1.4	2.5970 1.7844	0.8126
Log-logistic	IPI + DTIC PBO + DTIC	18.5 11.3	10.1 5.5	4.1 2.0	2.4 1.1	2.4079 1.6369	0.7709
Generalised-F	IPI + DTIC PBO + DTIC	20.8 13.1	12.5 7.7	5.8 3.8	3.6 2.5	2.8168 2.0924	0.7245
Log-normal	IPI + DTIC PBO + DTIC	19.5 12.3	10.2 5.5	3.3 1.4	1.5 0.5	2.2095 1.5385	0.6709
Spline - 1 knot - Odds	IPI + DTIC PBO + DTIC	21.0 12.8	12.5 7.0	4.6 2.9	2.1 1.7	2.4003 1.8644	0.5359
MWM	IPI + DTIC PBO + DTIC	21.2 13.4	11.3 8.1	1.5 2.1	0.1 0.5	1.9706 1.6170	0.3536

Figure 1: Reconstructed Kaplan-Meier plots of overall survival in CA184-024 and smoothed hazard plot for patients in CA184-024 (both data-cuts)
Key: DTIC, dacarbazine; IPI, ipilimumab; PBO, placebo.



Patients at risk

IPI + DTIC (original)	250	114	68	47	0	0	0
PBO + DTIC (original)	252	87	41	28	0	0	0
IPI + DTIC (updated)	250	116	68	50	43	41	41
PBO + DTIC (updated)	252	90	42	28	21	18	18

Figure 2: Overview of fitted models within observed period and extrapolation of survival.
Description: A: Short-term fit to observed data for ipilimumab + dacarbazine; B: Long-term extrapolation of survival for ipilimumab + dacarbazine; C: Short-term fit to observed data for placebo + dacarbazine; D: Long-term extrapolation of survival for placebo + dacarbazine.
Key: AJCC, American Joint Committee on Cancer; KM, Kaplan-Meier; MCM, mixture-cure model; MWM, mixture-Weibull model; ONS, Office for National Statistics; TA, Technology Appraisal.