ORIGINAL RESEARCH ARTICLE

A Case Study Examining the Usefulness of Cure Modelling for the Prediction of Survival Based on Data Maturity

Tim S. Grant[1] · Darren Burns[2] · Christopher Kiff[3] · Dawn Lee[2]

Abstract

Introduction Mixture modelling is increasingly being considered where a potential cure leads to a long life. Traditional methods use relative survival models for frail populations or cure models that have improper survival functions with theoretical infinite lifespans. Additionally, much of the work uses population data with long follow-up or theoretical data for method development.

Objective This case study uses life table data to create a proper survival function in a real-world clinical trial context. In particular, we discuss the impact of the length of trial follow-up on the accuracy of model estimation and the impact of extrapolation to capture long-term survival.

Methods A review of recent National Institute for Health and Clinical Excellence (NICE) immuno-oncological and chimeric antigen receptor (CAR) T-cell therapy submissions was performed to assess industry uptake and NICE acceptance of survival analysis methods incorporating the potential for long-term survivorship. The case study analysed a simulated trial-based dataset investigating a curative treatment with long-term mortality based on population life tables. The analysis examined three timepoints corresponding to early trial, end-of-trial follow-up and complete follow-up. Mixture modelling approaches were considered, including both cure modelling and relative survival approaches. The curves were evaluated based on the ability to estimate cure fractions and mean life in years within the time span the models are based on and when extrapolating to capture long-term behaviour. The survival curves were fitted with Weibull distributions using non-mixture and mixture cure models.

Results The performance of the cure modelling methods depended on the relative maturity of the data, indicating that care is needed when deciding when the methods should be applied. For progression-free survival, the cure fraction simulated was 15%. The cure fractions estimated using the traditional mixture cure model were 43% (95% confidence interval [CI] 30–57) at the first analysis time point (40 months), 15% (95% CI 12–20) at the end-of-study follow-up (153 months) and 0% (95% CI 0–100) at the end of follow-up. Other standard cure modelling methods produced similar results. For overall survival, we observed a similar pattern of goodness of fit, with a good fit for the end-of-study follow-up and poor fit for the other two data cuts. However, in this case, the estimate of the cure fraction was below the true value in the first analysis data.

Conclusions This case study suggests cure modelling works well with data in which the disease-specific events have had time to occur. Care is needed when extrapolating from immature data, and further information should support the estimation rather than relying on statistical estimates based on the trial alone.

✉ Tim S. Grant
tsgrant.dublin@gmail.com

1 BresMed Ireland Ltd, 28 River Gardens, Glasnevin, Dublin 9, Ireland

2 BresMed Health Solutions Ltd, Sheffield, UK

3 Bristol-Myers Squibb, Uxbridge, UK

1 Background

In recent years, pharmaceutical innovations in cancer treatments have led to global improvements in survival profiles, with long-term survivorship demonstrated across many indications [1]. However, developments in underlying clinical mechanisms (e.g. immuno-oncological and gene therapies) have meant that changes in survival outcomes affect more than simply the absolute level of survival observed [2–4]. An increasing number of current and future cancer drugs

Key Points for Decision Makers

Theoretical work in the development of mixture cure modelling often considers situations where events for cured and uncured patients will likely occur within similar timeframes; this is unrepresentative of most cases where cure modelling has been applied for economic modelling.

This case study suggests that cure modelling works well with data where disease-specific mortality has had enough time to occur but age-related mortality has not. In contrast, immature data give little information for estimating the probability of cure, potentially leading to overestimation or considerable uncertainty in the cure fraction.

As pivotal trial data alone are unlikely to be sufficient in a health technology appraisal setting, we recommend considering the use of supportive earlier-phase data, data on mechanism of action, biological data on the ability to translate observations across tumour types, and direct clinical input to assess the estimated cure fraction.

promise considerable increases in long-term survival for a proportion of treated patients. Consequently, these technologies are expected to affect not only the absolute level of survival observed in a treated cohort but also the long-term survival curve behaviour. This manifests as a plateau within the Kaplan–Meier (KM) curve, where observed events often appear to abruptly become less common, owing to the rising proportion of the surviving cohort that is at low risk of an event following treatment [5]. One example of this is the study by Schadendorf et al. [6] on ipilimumab in melanoma, in which there was a clear plateau in the KM curve. Other examples are available [7, 8].

For regulatory and health technology appraisal (HTA) bodies to evaluate the full clinical and economic value of a new technology, it is typically necessary to extrapolate beyond the observed clinical trial evidence. Many years of follow-up are required to track every patient in a cohort until their death or attrition, and HTA submissions usually take place years before these data are available. This is particularly true when a proportion of patients is at a lower level of risk, meaning a cohort could potentially require decades of follow-up to produce a complete set of survival observations. While methods to extrapolate beyond the limit of KM data are well-established and accepted by HTA agencies [9–12], these 'standard' methods are limited in their ability to accurately extrapolate more complicated survival functions. Consequently, there is an increasing need for modelling methods that capture long-term survivorship.

Cure modelling has existed in survival analysis since the 1950s [13]. The theoretical work behind mixture cure modelling (MCM) uses either a relative survival situation where it is likely that events—be it death or progression—for cured and uncured patients will occur within similar timeframes or the theoretical construct that, once cured, a subject lives forever.

In the relative survival setting, the mixture model uses a baseline survival function, $S^*(t)$, that captures the survival rate of general population of interest. The approach then models what fraction survive at that rate, π, and how to adjust that survival curve for the uncured fraction, $S'_U(t)$ as follows [14]:

$$S(t) = S^*(t)\left[\pi + (1 - \pi)S'_U(t)\right]$$

This formulation creates a term that is a product of survival functions, which has been shown to be valid if the hazard rates of the two survival functions are independent [15].

Cure modelling from a mixture model approach differs in that the first term, $S_C(t)$, which is the survival function for those cured, only applies to the cured population, and the second survival function, $S_U(t)$, only applies to the uncured population [16]:

$$S(t) = \pi S_C(t) + (1 - \pi)S_U(t)$$

In this way, $S_U(t)$ contains all causes of mortality as compared with $S'_U(t)$, which only contains the excess hazard of not being cured. In the theoretical construct of a true cure model, $S_C(t)$ is often 1, reducing to the improper survival function:

$$S(t) = \pi + (1 - \pi)S_U(t)$$

In a real-world context, relative survival models would be appropriate when either the disease is aggressive and in a much older population than commonly seen in clinical trials or the disease is indolent and in a younger population. These are settings where the shapes of the two survival curves are similar and it is the hazard rates that differ. The theoretical cure model with an improper survival function is not realistic in a health economics setting. Ultimately, implementing the models for health economic settings would require a combination of approaches that would use the parameters from the cure model to capture the uncured population and data from life tables or registries to capture the cured population.

Several papers have demonstrated successful application of cure modelling to longer-term, more representative datasets. For example, Mariotto et al. [17] investigated metastatic breast cancer with 8 years of follow-up, and Othus et al. [8] examined multiple myeloma at 15 years. Yu et al. [18] performed a simulation methodological study and identified that the bias of the methods depends on the maturity of the data.

Other approaches are in development that aim to model and extrapolate the survival behaviour of a heterogeneous or mixed population, composed of subjects from a cured population and an uncured population, from intermediate outcomes or capture an overall shape that does not fit the traditional parametric approaches without formally defining a mix of patients (outside the scope of this article). These include response-based and spline-based approaches, competing risk approaches and simply using the KM estimates directly until a chosen timepoint and basing extrapolations on the remaining observations [19–23].

In this study, we present the findings from a literature review exploring current attitudes of UK decision makers towards using MCM or spline-based methods. We then report on the ability of 'standard' and cure methods to capture and extrapolate survival of a mixed population using a simulated case study dataset based on actual disease progression rates from the literature, coupled with the conditions of a clinical trial from a similar population. We investigated the impact of introducing a hypothetical therapy with the potential for long-term survivorship into this simulated population, testing the ability of each potential method to fit the underlying data at three different durations of follow-up. The results are then discussed across timepoints within the three types of comparisons: estimation, apparent accuracy of the fits and ability to extrapolate. Finally, we highlight the need for guidance and provide evidence to aid development of such guidance.

2 Methods

2.1 National Institute for Care and Excellence (NICE) Submission Review

Information was extracted from the final appraisal document, committee papers, Evidence Review Group report and appraisal consultation document for immuno-oncological and chimeric antigen receptor (CAR) T-cell therapy submissions to the National Institute for Health and Care Excellence (NICE) published between 2017 and 2018. Data extraction focused on the survival analysis methods used and how they were received and the final decision on the appropriate method.

2.2 Analysis Methods

The analysis methods considered for estimation were non-mixture cure models (NMCMs) and MCMs. Only the Weibull distribution was considered [24]. The implementations and theories behind them are presented here.

2.2.1 Non-mixture and Mixture Cure Models

The NMCM and MCM were implemented with the flexsurvcure package in R and distinguished with the mixture parameter being FALSE or TRUE, respectively.

In this context, the NMCM survival function of the heterogeneous population is of the form:

$$S(t) = \pi^{F_Z(t)} = \exp\big(\ln(\pi) - \ln(\pi)S_Z(t)\big)$$

where $S(t)$ is the population survival function, π is the cure fraction, $F_z(t)$ is the cumulative distribution function and $S_Z(t)$ is the survival function for the uncured subjects. For the Weibull distribution, the software reported a cure fraction, shape parameter and rate parameter for the model.

The MCM is a modification of the standard mixture and can be expressed as:

$$S(t) = \pi S_C(t) + (1 - \pi)S_U(t)$$

where $S_C(t)$ is the survival function for the cured population and $S_U(t)$ is the survival distribution for the uncured population. When run as a cure model, $S_C(t)$ was set to unity and π, and the parameters for $S_U(t)$ were estimated with 95% confidence intervals.

2.2.2 Relative Survival

Where cure models assume a survival function of 1 for those cured, relative survival models assume a survival function for those cured and a second survival function for those who remain uncured. The survival function for those who are uncured is a relative survival function that captures the difference between the hazard rates, which when independent is a product term. The model as presented in Lambert [25] is generally expressed as:

$$S(t) = S_C(t)\big(\pi + (1 - \pi)S'_U(t)\big)$$

In this formulation, the survival curve for cured subjects is S_C and for uncured subjects is $S_C S'_U(t)$. Whether the cure model is a relative survival model or not is controlled by the bhazard option in flexsurvcure. When the option is not used, a traditional cure model with a cured survival function of 1 is estimated. If the bhazard option is used, the model is a relative survival model. The baseline hazards (the rates for the survival curve for the cured subjects) are constants and can be either constant across all subjects or vary by subject. Both have been considered identified as constant baseline and individual baseline. In this model, $S_C(t)$ is known and the cure fraction, π, and the parameters for $S'_U(t)$ are estimated with 95% confidence intervals.

In this work, all three options, no baseline, constant baseline and individual baseline were considered.

2.3 Implementation of Cure Models

The cure models were implemented based on the work of Lambert et al. [14, 25], which adds the lifetime mortality survival curve to the cure models. Essentially, the parameters are estimated assuming either cure or a selected baseline hazard rate. The baseline survival function, $S_C(t)$, is then replaced with information from the life tables in a non-parametric way, $S^*(t)$. The model for the NMCM is:

$$S(t) = S^*(t)\exp\big(\ln(\pi) - \ln(\pi)S_Z(t)\big)$$

where $S^*(t)$ is the expected survival function, which, in this case, came from general population mortality life tables. Note, $S_C(t)$ is 1 in Sect. 2.2 and omitted in that case. Similarly, for the MCM, this can be expressed as:

$$S(t) = S^*(t)\big(\pi + (1 - \pi)S_U(t)\big)$$

2.4 Simulated Dataset

A dataset of times to progression and times to death was simulated based on a hypothetical single-arm clinical trial design. Recruitment dates for each subject and the potential for withdrawing from the study were included in the simulation. Dataset characteristics were selected to reflect real life as much as possible (i.e. using non-round numbers). Survival times of non-cured subjects were taken from a published paper on an aggressive disease with a high hazard rate to best reflect the type of data likely to be seen in a real-life scenario [26]. The dataset comprised 347 patients. These patients were subjected to an intervention that, if successful, would allow their lifetime risks to return to those of the general population. If unsuccessful, progression would generally occur before 100 months, and death would usually occur before 150 months.

Each subject was then identified as belonging to one of three groups at random: those who remained uncured, those cured after progression and those cured before progression. Hypothetical cure fractions were defined within the simulated dataset as 15% pre-progression and 25% post-progression, for an overall cure rate of 40%. These cure fractions allowed for examining a higher cure rate (40%) and a lower cure rate of 15% and were not intended to be accurate to a specific therapy. It was considered important to allow for long-term survival post-progression because this has been observed in previous immunotherapy trials. Transition matrices and details of how these were derived are presented in the electronic supplementary material (ESM)-1.

Study characteristics were also addressed in the simulation, including time of recruitment, loss to follow-up and planned study analysis time points. Recruitment was assumed to occur over a 27-month period. Subject start times were assumed to be uniformly distributed over this period. Loss to follow-up due to subjects discontinuing from the trial either because of adverse events, drop-out or withdrawal of consent was simulated by selecting and censoring 18 subjects at random pre-progression and an additional 36 subjects post-progression, representing a 15% loss to follow-up. The time of discontinuation was simulated uniformly across the time period of withdrawal. Three study analysis time points (data cuts) were considered for the simulation:

> Time of first analysis: 40 months from the start of recruitment. Given that most UK submissions conducted for immuno-oncology therapies occur with less mature data cuts, this represented a reasonable maximum maturity level for a first product submission.
> End of study with follow-up: 153 months from the start of recruitment. This was the latest an observation could be recorded, according to the simulated study design. This is considerably longer than the follow-up available in most NICE submissions.
> Complete dataset: 500 months of follow-up. This does not reflect data collected in clinical trials and was possible only because these data were simulated. This set provided a comparison for the extrapolated results of models from the prior datasets.

The simulated complete dataset KM curves are given in ESM 1. The key value this work adds is considering how a real-life data-collection process affects survival-type data and provides an example of how the incorrect application of cure modelling methods can lead to erroneous conclusions.

2.5 Evaluation

The various methods were evaluated based on their ability to estimate the cure fraction, the apparent goodness of fit to the dataset the models were based on and the ability of each model to extrapolate to the true complete dataset. The ability to estimate the cure fraction, the shape parameter and the scale parameter was based on 95% confidence intervals as reported by the software package. These confidence intervals were compared with values from fits directly to the uncured population, which could be obtained because the cured status was known from the simulation process. The ability of the methods to fit curves up to the timepoint in question was based on comparisons with the censored KM curves for each data cut. The ability of the models to extrapolate to the true curve were based on comparisons with the complete dataset.

3 Results

3.1 NICE Submission Review

Summaries of the survival analysis in all considered appraisals are provided in ESM 2, along with a more detailed review. In total, 23 completed NICE appraisals of immuno-oncological interventions published between 2017 and 2018 were identified. Of these, 20 reported at least some criticism relating to immaturity of data, with the majority having a minimum overall survival (OS) follow-up of ≤ 12 months (11/18 that reported minimum follow-up). Furthermore, eight submissions submitted OS data with ≤ 24 months of maximum follow-up (see Tables 1 and 2 in ESM 2).

Four completed submissions (technology appraisal [TA] 478, TA492, TA520 and TA525) included the presentation of MCM in some form [27–30]. One submission (TA417) presented a method assuming all 5-year survivors remaining on treatment had the same mortality as the general population, thereby assuming a cure fraction not based on trial data. The other 18 submissions considered a mixture of 'standard' parametric survival models, spline models, fractional polynomial models (all applied to network meta-analyses) and piecewise approaches using KM data directly or switching to using general population mortality at some trigger point. Of the five submissions presenting various cure methods, no reimbursement decisions were explicitly made using MCM results with any cure fraction > 0%, although for TA417 the final appraisal document did state that NICE was willing to consider some scenarios assuming a cure fraction, just not as the decision-making result.

To supplement this review, all available materials (at the time of writing) from the three NICE appraisals for CAR-T therapies (ID1166 for tisagenlecleucel-T; ID1115 for axi-cabtagene ciloleucel; ID1167 for tisagenlecleucel-T) were also reviewed. These appraisals were important to review because they consider the latest group of therapies to demonstrate plateaus within their KM curves, and the mode of action of CAR-T therapies suggests candidacy for an MCM approach. In all three cases, MCM was accepted in some form by NICE and the respective Evidence Review Groups.

Overall, at this time, no completed HTA in the UK has supported survival extrapolations estimated solely using trial-based MCM (with a non-zero cure fraction), but the method has received consideration in ongoing appraisals that show clear evidence of plateau. However, there appears to be no clear consensus on the appropriateness of MCM in different survival analysis contexts.

3.2 Estimation Results

The estimation results for progression-free survival (PFS; Table 1) indicated good estimation for the cure modelling approaches at 153 months; however, estimated values were above the true value for all methods at 40 months and below the true value for all methods at 500 months. The simulation provided a true value of 15% cure for these estimates, whereas the cure modelling methods estimated just over 40% at 40 months and 0% at 500 months.

The estimation results for OS (Table 2) at 153 months were still reasonable for the cure modelling methods, with the true value of 40% falling within the confidence intervals. The cure fractions were below the true value for both 40 months and the complete dataset, with an estimate of 0% cured.

The impact of cure modelling on the estimated mean life-years (shown in Table 3) demonstrated that estimation of the time to progression from data at 40 months exaggerated the efficacy of the treatment. Estimation based on the complete dataset understated the efficacy of the treatment in terms of progression.

Mean life-years for OS (Table 4) were below the true values for the data cut at 40 months and the complete dataset for the cure modelling approaches. At 153 months, the approaches provided reasonable estimates.

Table 1 Estimation of the cure fraction (PFS)

	PFS percentage cured (95% CI)		
	Month 40	Month 153	Complete FU
True value	0.15	0.15	0.15
Non-mixture cure modelling	0.42 (0.28–0.57)	0.15 (0.1–0.2)	0 (0–0.09)
Non-mixture cure modelling—constant baseline	0.43 (0.3–0.58)	0.16 (0.11–0.22)	0 (0–0.09)
Non-mixture cure modelling—individual baseline	0.44 (0.3–0.58)	0.16 (0.11–0.22)	0 (0–0.09)
Mixture cure modelling	0.43 (0.3–0.57)	0.15 (0.12–0.2)	0 (0–1)
Mixture cure modelling—constant baseline	0.44 (0.32–0.58)	0.17 (0.13–0.21)	0 (0–0.97)
Mixture cure modelling—individual baseline	0.45 (0.32–0.58)	0.17 (0.13–0.21)	0 (0–0.98)

CI confidence interval, *FU* follow-up, *PFS* progression-free survival

△ Adis

Table 2 Estimation of the cure fraction (OS)

	OS percentage cured (95% CI)		
	Month 40	Month 153	Complete FU
True value	0.40	0.40	0.40
Non-mixture cure model	0 (0–1)	0.43 (0.36–0.49)	0 (0–1)
Non-mixture cure model—constant baseline	0 (0–1)	0.47 (0.4–0.54)	0 (0–1)
Non-mixture cure model—individual baseline	0 (0–1)	0.47 (0.4–0.54)	0 (0–1)
Mixture cure model	0 (0–1)	0.43 (0.37–0.49)	0 (0–0.99)
Mixture cure model—constant baseline	0 (0–1)	0.47 (0.41–0.54)	0 (0–0.95)
Mixture cure model—individual baseline	0.01 (0–1)	0.47 (0.41–0.54)	0 (0–0.94)

CI confidence interval, *FU* follow-up, *OS* overall survival

Table 3 Estimation of the mean life-years (PFS)

	PFS mean life-years (95% CI)		
	Month 40	Month 153	Complete FU
True value	6.06	6.06	6.06
Non-mixture cure model	4.5 (3.39–5.77)	4.66 (3.98–5.42)	5.23 (4.60–6.00)
Non-mixture cure model—constant baseline	11.62 (8.75–14.99)	5.98 (4.99–7.20)	5.05 (4.46–7.47)
Non-mixture cure model—individual baseline	12.02 (9.09–15.34)	6.23 (5.21–7.56)	5.27 (4.64–7.67)
Mixture cure model	12.06 (9.32–15.36)	6.23 (5.2–7.61)	5.28 (4.65–7.29)
Mixture cure model—constant baseline	11.86 (9.21–14.94)	5.98 (5.06–7.14)	5.23 (4.65–25.71)
Mixture cure model—individual baseline	12.24 (9.48–15.42)	6.22 (5.18–7.41)	5.46 (4.79–25.06)

CI confidence interval, *FU* follow-up, *PFS* progression-free survival

Table 4 Estimation of the mean life-years (OS)

	OS mean life-years (95% CI)		
	Month 40	Month 153	Complete FU
True value	14.07	14.07	14.07
Weibull survival model	6.42 (4.48–9.02)	10.96 (9.62–12.25)	11.31 (10.35–12.30)
Non-mixture cure model	6.66 (5.75–25.71)	13.47 (12.09–14.97)	11.31 (10.72–25.71)
Non-mixture cure model—constant baseline	6.66 (5.54–25.71)	14.28 (12.79–15.83)	12 (11.53–25.71)
Non-mixture cure model—individual baseline	6.68 (5.95–25.71)	14.3 (12.79–15.83)	12.03 (11.42–25.71)
Mixture cure model	6.47 (5.03–25.71)	13.53 (12.27–14.81)	11.31 (10.41–25.47)
Mixture cure model—constant baseline	6.51 (4.81–25.71)	14.32 (13.06–15.75)	12.01 (11.05–24.58)
Mixture cure model—individual baseline	6.53 (4.82–25.71)	14.34 (12.96–15.74)	12.03 (11.09–25.32)

The upper bound of 25.71 years represents mean general population survival, i.e. a cure fraction of 100%

CI confidence interval, *FU* follow-up, *OS* overall survival

3.3 Survival Curves

The estimated survival curves for the three timepoints for OS (Figs. 1, 2, 3) demonstrated that the curves generally fitted well over the timespan the model is based on. However, with the exception of the end-of-study timepoint, the curves failed to capture the true survival curves across time.

For the interim analyses timepoint, the curves captured the true curve up to 40 months. At 40 months, the time of the data cut, the cure modelling approaches continued to capture the behaviour up to 100 months and then fell below the true survival curve.

For the end-of-study timepoint, the methods all followed the true survival curve. The cure modelling approaches fell slightly below observed survival from approximately 100 months up to 250 months and were slightly above observed survival from approximately 300–400 months.

The models based on the complete dataset provided curves that generally fell below observed survival, particularly beyond 150 months. The cure modelling approaches
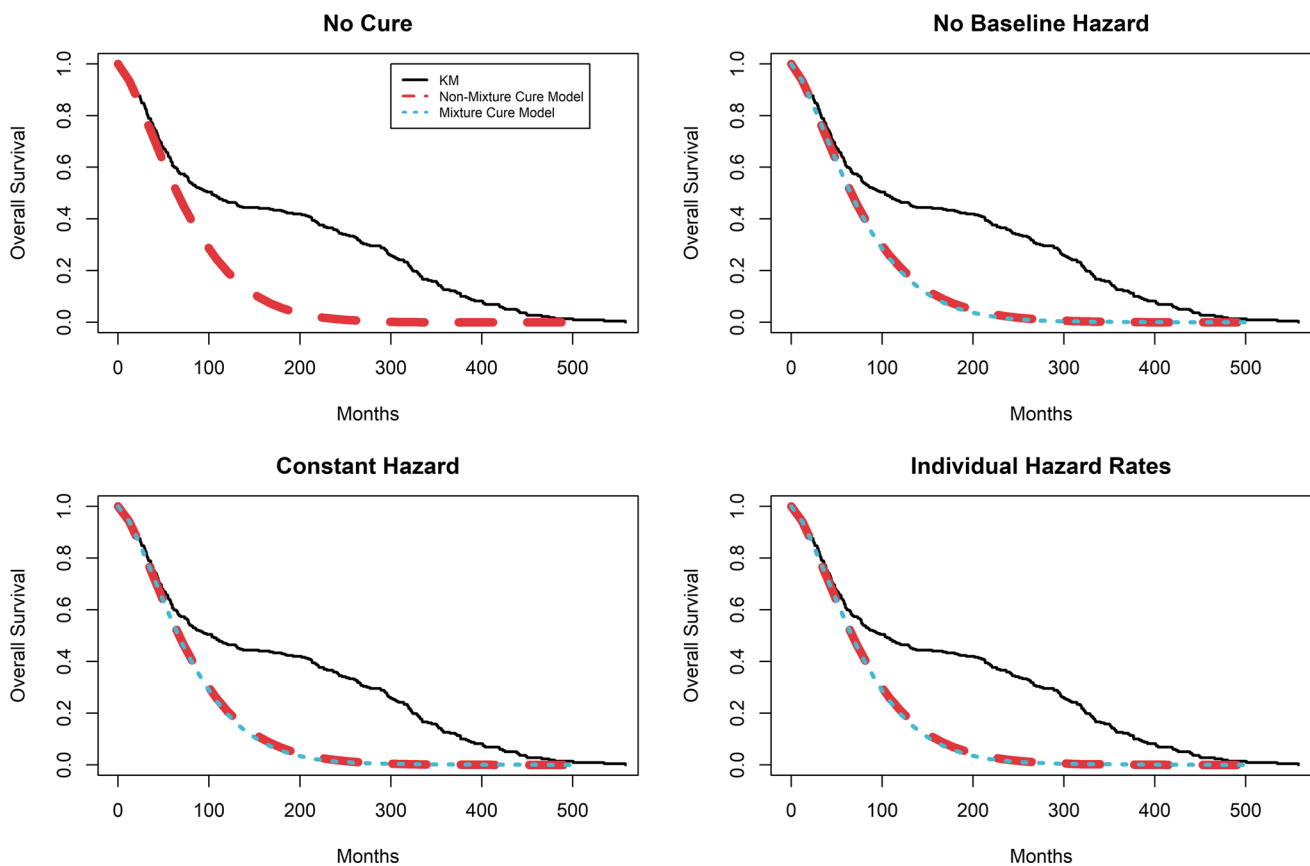
Fig. 1 Fit of the models to the overall survival KM data: interim analyses (month 40). Plots of fit to survival data for KM, non-mixture cure model, and mixture cure model for **a** no cure, **b** no baseline hazard, **c** constant hazard and **d** individual hazard rates. *KM* Kaplan–Meier. Figure created in R statistical software

rose above observed survival from approximately 50–150 months before falling well below the true curve.

Similar to OS, for PFS there was a good fit to survival data using the end-of-study dataset, and poor fit when using the other two datasets (see ESM 3).

4 Discussion

Despite the mode of action of newer oncology therapies, there is often no visible plateau in KM data because of the short follow-up available at the time of HTA submission to NICE (most have < 12 months of minimum follow-up and many have a maximum follow-up of < 24 months). Consequently, MCM is often difficult to justify to most reviewers, even in situations where it would add value. Our results provide evidence to a sceptical reviewer that extrapolation methods other than the standard methods suggested in NICE Technical Support Document 14 [12] may be appropriate in some cases and improve upon the standard methods. Additionally, we provide recommendations on identifying these situations.

Overall, MCM appeared to be the least accepted of all the parametric survival approaches, despite its obvious relevance to therapies with the potential for long-term survivorship. Only one of the four reviewed submissions rejected an MCM approach explicitly because of a lack of evidence of a survival plateau; other MCMs were rejected because of a lack of justification for the estimated cure fractions for PFS and OS or because of the similarity of the final resulting extrapolation to methods with more precedent, such as KM plus extrapolation. However, when considering NICE submissions in CAR-T therapies, MCM appeared to be better received, being accepted but not endorsed in every case so far.

Guidance on when MCM is likely to be suitable would assist committees in confidently selecting this method when it provides the most reasonable extrapolation of long-term survival, for example, when the mode of action or clinical expectation are suggestive of a future survival plateau.

The primary result of this case study is the evidence supporting the use of the method with data that have reached an appropriate degree of maturity. Further to this, caution is needed when the techniques are applied to immature data,
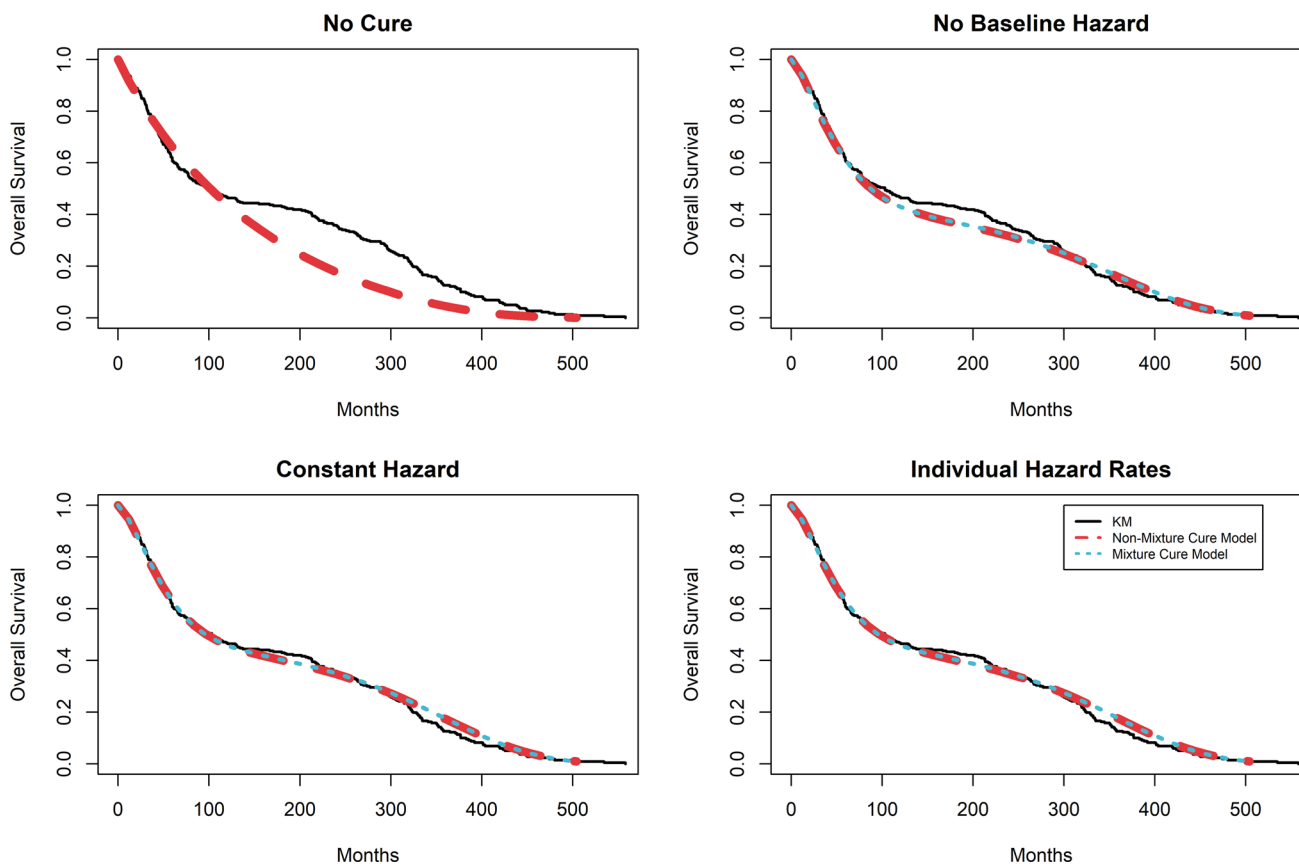
Fig. 2 Fit of the models to the overall survival data: end-of-study analyses (month 153). Plots of fit to survival data for KM, non-mixture cure model and mixture cure model for **a** no cure, **b** no baseline hazard, **c** constant hazard and **d** individual hazard rates. *KM* Kaplan–Meier. Figure created in R statistical software

resulting in considerable inaccuracy in the cure fraction. For the interim analysis results—which, if successful, would have concluded the trial and resulted in reimbursement—the models did not extrapolate well, largely because of poor estimation of the cure fraction for PFS and for OS. The PFS estimation being above the true value was likely due to the high concentration of censored observations at or just before the 40-month point, which was when the data collection was simulated to occur. Up to 40 months, all of the models performed adequately, but there was no indication of superiority to standard approaches. For OS, as the change in hazards had not yet been observed, the cure models failed to predict the impending plateau. Estimation of the mean life-years demonstrated the high risk of using these methods on immature data. In this example, PFS was estimated to be longer than the OS because of the PFS overestimation.

For the end-of-study analyses, the methods worked well on data of this maturity; while the data would be considered incomplete given the missing long-term follow-up results, mortality due to the disease was largely completed. The cure modelling approaches improved the estimation of the extrapolated cure model. The NMCM outperformed the MCMs

based on visual inspection. While there was a concentration of censored points at or just before 153 months in these models, these were generally all-cause events not associated with the disease in question and so did not inflate the cure fraction estimate.

The methods did not work well when applied to complete datasets, as all subjects had an event except those lost from the study either because of withdrawal or adverse effects. This created a situation where the cure fraction was based on those lost from the study and not the plateau in the middle of the survival curve. As a result, the cure fraction estimate fell below the true value, and the resulting curve was a single Weibull curve to the complete dataset instead of a Weibull curve for those events related to the disease. Non-parametric methods such as a KM approach might be preferred with complete datasets and alternative methods such as spline modelling and competing risk approaches. Separating the modelling of disease-related and non-disease-related mortality within the patient-level data may also be possible with more mature data where both types of events are observed.

A serious limitation to these approaches is the estimation methods in the software packages at this time. A constant
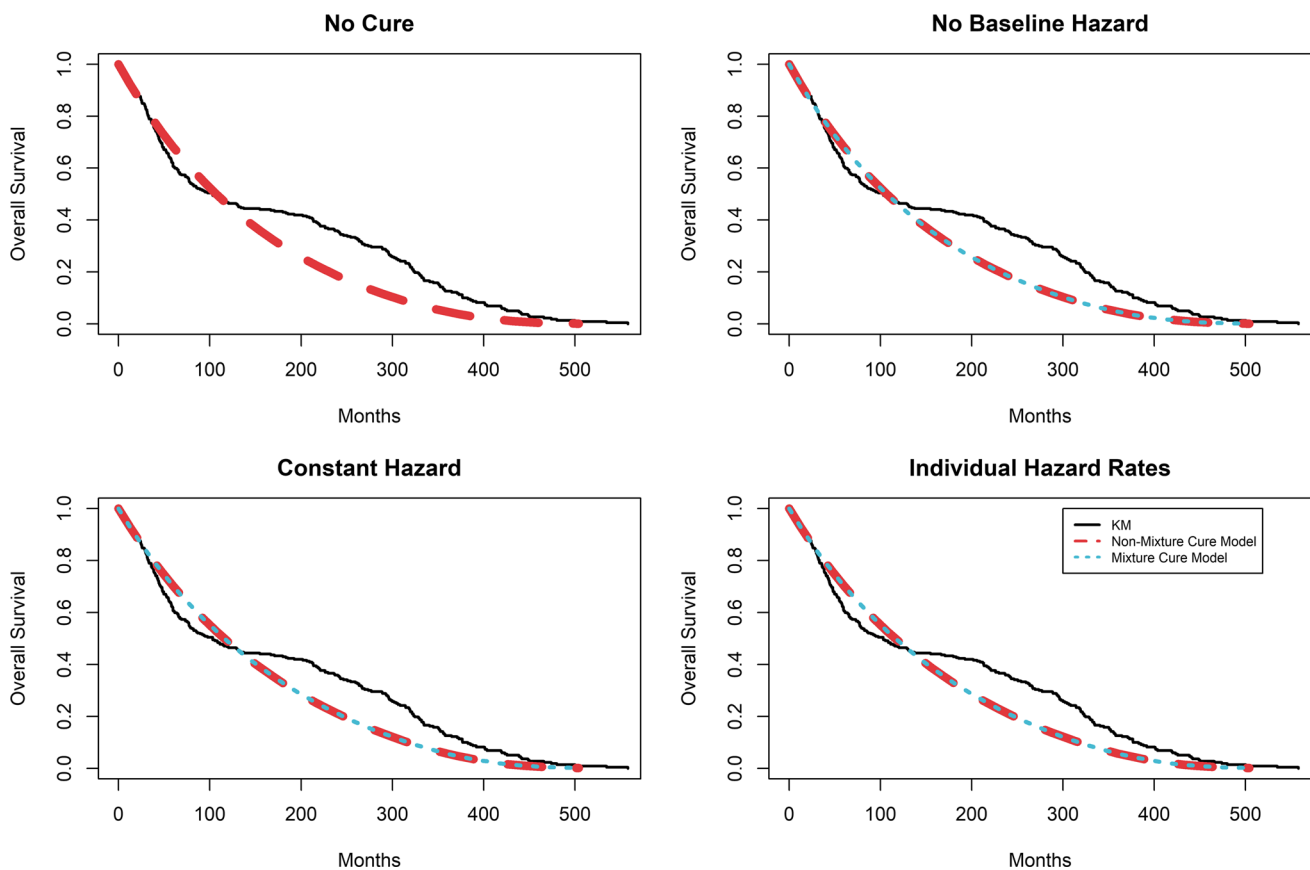
Fig. 3 Fit of the models to the overall survival data: complete dataset. Plots of fit to survival data for KM, non-mixture cure model and mixture cure model for **a** no cure, **b** no baseline hazard, **c** constant hazard and **d** individual hazard rates. *KM* Kaplan–Meier. Figure created in R statistical software

hazard rate, at the individual or population level, for the baseline hazard is unrealistic in these situations. Relative survival models are more appropriate when the cured population is still at high risk, which is not the situation being examined here. However, little difference between relative survival and cure models was observed in this case study.

Previous work on MCM has focused largely on the datasets with long-term follow-up (over 5 years) [2, 31]. Our findings align with these examples, providing an ideal basis for cure modelling, with available data that show the plateau before general population mortality has started to gain influence on the survival of the remaining cohort.

5 Conclusions

Healthcare decision makers have a difficult task in deciding which therapies to recommend; they must balance the need to use resources wisely with the need to treat life-threatening diseases in patients who are unlikely to have time to wait for more data to be generated. Decision makers must therefore

identify an acceptable way to use the often short-term follow-up data to make long-term survival predictions.

In this case study with simulated data, we demonstrated that cure modelling approaches can be useful with appropriate data but will be misleading if the data are either immature or very mature; they should therefore be used with care. Within this simulation, cure modelling worked best when most disease-related events had occurred and most other-cause mortality events were yet to occur. As this will vary by disease, clinical input will be important in the application of cure modelling methods. For example, the metastatic breast cancer work of Mariotto et al. [17] would indicate requiring at least 10 years of follow-up, whereas Othus et al. [8] used a 5-year follow-up, supporting the use of cure modelling for economic evaluation with this dataset. The simulated dataset from this study suggests that 10 years of data are required in this case. Secondary to the rate at which disease events occur is the average age of patients in the study. If the patients are mostly elderly, then all-cause mortality will begin sooner, leading to a different environment than the case examined here.

While this is a case study approach, and the conclusions are specific to these data, it makes the case that caution must be used with the application of cure modelling approaches and that simulation studies are required to better understand the limitations of the methods. Additional studies could establish how common these findings are or investigate specific aspects of the models. For example, the baseline hazard rates are relatively low, representing a healthy population. For diseases that have high mortality rates even when patients are cured, these aspects might yield greater differences. This will be particularly interesting in the application of individual baseline hazard rates, as the distribution of patients by sex and age might influence when this additional consideration is required. Sourcing mortality data for cured patients may also prove difficult, particularly in situations where it is not appropriate to use general population mortality.

Based on this case study, for situations with potential for cure but incomplete survival data, we would first recommend inspection of the KM curve for the presence of a plateau to determine whether the cure modelling output can provide reasonable extrapolations for predicting long-term events. This is consistent with the recommendation from Amico and Van Keilegom [32], who also referenced a test for a plateau proposed by Maller and Zhou [33]. Second, pivotal trial data alone are unlikely to provide sufficient information to estimate cure fractions in the HTA setting. We recommend using supportive earlier-phase data, data on mechanism of action and biological data on the ability to translate observations across tumour types and direct clinical input in assessing the estimated cure fraction. We suggest that a promising area for future research would be methods estimating the cure fraction that synthesise all available data and also incorporate direct expert elicitation.

Compliance with Ethical Standards

References

1. Allemani C, Matsuda T, Di Carlo V, Harewood R, Matz M, Niksic M, et al. Global surveillance of trends in cancer survival 2000-14 (CONCORD-3): analysis of individual records for 37 513 025 patients diagnosed with one of 18 cancers from 322 population-based registries in 71 countries. Lancet (Lond Engl). 2018;391(10125):1023–75. https://doi.org/10.1016/s0140-6736(17)33326-3.

2. Bullement A, Meng Y, Cooper M, Lee D, Harding TL, O'Regan C et al. A review and validation of overall survival extrapolation in health technology assessments of cancer immunotherapy by the National Institute for Health and Care Excellence: how did the initial best estimate compare to trial data subsequently made available? J Med Econ. 2018;2018:1–10. https://doi.org/10.1080/13696998.2018.1547303.

3. Dine J, Gordon R, Shames Y, Kasler MK, Barton-Burke M. Immune checkpoint inhibitors: an innovation in immunotherapy for the treatment and management of patients with cancer. Asia-Pac J Oncol Nurs. 2017;4(2):127–35. https://doi.org/10.4103/apjon.apjon_4_17.

4. Latimer N. Estimating survival benefit for health technology assessment: new challenges presented by immuno-oncology treatments? BBS/PSI 1-Day Scientific Meeting: Empower the immune system to fight cancer. Basel: Switzerl. 2017.

5. Kaplan EL, Meier P. Nonparametric estimation from incomplete observations. J Am Stat Assoc. 1958;53(282):457–81. https://doi.org/10.2307/2281868.

6. Schadendorf D, Hodi FS, Robert C, Weber JS, Margolin K, Hamid O, 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(17):1889–94. https://doi.org/10.1200/jco.2014.56.2736.

7. Neelapu SS, Locke FL, Bartlett NL, Lekakis LJ, Miklos DB, Jacobson CA, et al. Axicabtagene ciloleucel CAR T-cell therapy in refractory large B-cell lymphoma. New Engl J Med. 2017;377(26):2531–44. https://doi.org/10.1056/NEJMoa1707447.

8. Othus M, Barlogie B, LeBlanc ML, Crowley JJ. Cure models as a useful statistical tool for analyzing survival. Clin Cancer Res. 2012;18(14):3731–6. https://doi.org/10.1158/1078-0432.ccr-11-2859.

9. Canadian Agency for Drugs and Technologies in Health. Guidelines for the Economic Evaluation of Health Technologies: Canada—4th Edition. 2017. https://www.cadth.ca/about-cadth/how-we-do-it/methods-and-guidelines/guidelines-for-the-economic-evaluation-of-health-technologies-canada. Accessed 31 Jan 2019.

10. Haute Autorité de Santé. Choices in Methods for Economic Evaluation. 2012. https://www.has-sante.fr/portail/upload/docs/application/pdf/2012-10/choices_in_methods_for_economic_evaluation.pdf. Accessed 31 Jan 2019.

11. Jackson C. flexsurv: a platform for parametric survival modeling in R. J Stat Softw. 2016;70(8):33. https://doi.org/10.18637/jss.v070.i08.

12. Latimer N. NICE DSU Technical Support Document 14: survival analysis for economic evaluations alongside clinical trials—extrapolation with patient-level data. 2011. http://nicedsu.org.uk/wp-content/uploads/2016/03/NICE-DSU-TSD-Survival-analysis.updated-March-2013.v2.pdf. Accessed 31 Jan 2019.

13. Berkson J, Gage RP. Survival curve for cancer patients following treatment. J Am Stat Assoc. 1952;47(259):501–15. https://doi.org/10.2307/2281318.

14. Lambert PC, Thompson JR, Weston CL, Dickman PW. Estimating and modeling the cure fraction in population-based cancer survival analysis. Biostat (Oxf, Engl). 2007;8(3):576–94. https://doi.org/10.1093/biostatistics/kxl030.

15. De Angelis R, Capocaccia R, Hakulinen T, Soderman B, Verdecchia A. Mixture models for cancer survival analysis: application to population-based data with covariates. Stat Med. 1999;18(4):441–54. https://doi.org/10.1002/(sici)1097-0258(19990228)18:4%3c441:aid-sim23%3e3.0.co;2-m.

16. McLachlan Geoffrey J, Peel D. Finite mixture models. New York: Wiley; 2000.

17. Mariotto AB, Etzioni R, Hurlbert M, Penberthy L, Mayer M. Estimation of the number of women living with metastatic breast cancer in the United States. Cancer Epidemiol Biomark Prev. 2017. https://doi.org/10.1158/1055-9965.epi-16-0889.

18. Yu B, Tiwari RC, Cronin KA, Feuer EJ. Cure fraction estimation from the mixture cure models for grouped survival data. Stat Med. 2004;23(11):1733–47. https://doi.org/10.1002/sim.1774.

19. Bolard P, Quantin C, Abrahamowicz M, Esteve J, Giorgi R, Chadha-Boreham H, et al. Assessing time-by-covariate interactions in relative survival models using restrictive cubic spline functions. J Cancer Epidemiol Prev. 2002;7(3):113–22.

20. Giorgi R, Abrahamowicz M, Quantin C, Bolard P, Esteve J, Gouvernet J, et al. A relative survival regression model using B-spline functions to model non-proportional hazards. Stat Med. 2003;22(17):2767–84. https://doi.org/10.1002/sim.1484.

21. 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):2175–97. https://doi.org/10.1002/sim.1203.

22. Gibson E, Koblbauer I, Begum N, Dranitsaris G, Liew D, McEwan P, et al. Modelling the survival outcomes of immuno-oncology drugs in economic evaluations: a systematic approach to data analysis and extrapolation. PharmacoEconomics. 2017;35(12):1257–70. https://doi.org/10.1007/s40273-017-0558-5.

23. Ouwens MJNM, Mukhopadhyay P, Zhang Y, Huang M, Latimer N, Briggs A. Estimating lifetime benefits associated with immuno-oncology therapies: challenges and approaches for overall survival extrapolations. PharmacoEconomics. 2019;37(9):1129–38. https://doi.org/10.1007/s40273-019-00806-4.

24. Andersson TM, Dickman PW, Eloranta S, Lambert PC. Estimating and modelling cure in population-based cancer studies within the framework of flexible parametric survival models. BMC Med Res Methodol. 2011;11:96. https://doi.org/10.1186/1471-2288-11-96.

25. Lambert PC. Modeling of the cure fraction in survival studies. Stata J. 2007;7(3):351–75.

26. Weller M, Felsberg J, Hartmann C, Berger H, Steinbach JP, Schramm J, et al. Molecular predictors of progression-free and overall survival in patients with newly diagnosed glioblastoma: a prospective translational study of the German Glioma Network. J Clin Oncol. 2009;27(34):5743–50. https://doi.org/10.1200/jco.2009.23.0805.

27. National Institute for Health and Care Excellence. Brentuximab vedotin for treating relapsed or refractory systemic anaplastic large cell lymphoma [TA478]. 2017. https://www.nice.org.uk/guidance/ta478. Accessed 31 Jan 2019.

28. National Institute for Health and Care Excellence. Atezolizumab for untreated PD-L1-positive locally advanced or metastatic urothelial cancer when cisplatin is unsuitable [TA492]. 2017. https://www.nice.org.uk/guidance/ta492. Accessed 31 Jan 2019.

29. National Institute for Health and Care Excellence. Atezolizumab for treating locally advanced or metastatic non-small-cell lung cancer after chemotherapy [TA520]. 2018. https://www.nice.org.uk/guidance/ta520. Accessed 31 January 2019.

30. National Institute for Health and Care Excellence. Atezolizumab for treating locally advanced or metastatic urothelial carcinoma after platinum-containing chemotherapy [TA525]. 2018. https://www.nice.org.uk/guidance/ta525. Accessed 31 Jan 2019.

31. Othus M, Bansal A, Koepl L, Wagner S, Ramsey S. Accounting for cured patients in cost-effectiveness analysis. Value Health. 2017;20(4):705–9. https://doi.org/10.1016/j.jval.2016.04.011.

32. Amico M, Van Keilegom I. Cure models in survival analysis. Annu Rev Stat Appl. 2018;5(1):311–42. https://doi.org/10.1146/annurev-statistics-031017-100101.

33. Maller RA, Zhou X. Survival analysis with long-term survivors. New York: Wiley; 1996.