



This is a repository copy of *Mixture and non-mixture cure models for health technology assessment: what you need to know*.

White Rose Research Online URL for this paper:  
<https://eprints.whiterose.ac.uk/215602/>

Version: Published Version

---

**Article:**

Latimer, N.R. [orcid.org/0000-0001-5304-5585](https://orcid.org/0000-0001-5304-5585) and Rutherford, M.J. [orcid.org/0000-0003-1557-6697](https://orcid.org/0000-0003-1557-6697) (2024) *Mixture and non-mixture cure models for health technology assessment: what you need to know*. *PharmacoEconomics*, 42. pp. 1073-1090. ISSN 1170-7690

<https://doi.org/10.1007/s40273-024-01406-7>

---

**Reuse**

This article is distributed under the terms of the Creative Commons Attribution-NonCommercial (CC BY-NC) licence. This licence allows you to remix, tweak, and build upon this work non-commercially, and any new works must also acknowledge the authors and be non-commercial. You don't have to license any derivative works on the same terms. More information and the full terms of the licence here:  
<https://creativecommons.org/licenses/>

**Takedown**

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing [eprints@whiterose.ac.uk](mailto:eprints@whiterose.ac.uk) including the URL of the record and the reason for the withdrawal request.



[eprints@whiterose.ac.uk](mailto:eprints@whiterose.ac.uk)  
<https://eprints.whiterose.ac.uk/>



# Mixture and Non-mixture Cure Models for Health Technology Assessment: What You Need to Know

Nicholas R. Latimer<sup>1,2</sup> · Mark J. Rutherford<sup>3</sup>

Accepted: 4 June 2024  
© The Author(s) 2024

## Abstract

There is increasing interest in the use of cure modelling to inform health technology assessment (HTA) due to the development of new treatments that appear to offer the potential for cure in some patients. However, cure models are often not included in evidence dossiers submitted to HTA agencies, and they are relatively rarely relied upon to inform decision-making. This is likely due to a lack of understanding of how cure models work, what they assume, and how reliable they are. In this tutorial we explain why and when cure models may be useful for HTA, describe the key characteristics of mixture and non-mixture cure models, and demonstrate their use in a range of scenarios, providing Stata code. We highlight key issues that must be taken into account by analysts when fitting these models and by reviewers and decision-makers when interpreting their predictions. In particular, we note that flexible parametric non-mixture cure models have not been used in HTA, but they offer advantages that make them well suited to an HTA context when a cure assumption is valid but follow-up is limited.

## Key Points for Decision Makers

Many new treatments in a range of disease areas offer the potential for cure in some patients.

Cure models are not commonly used to inform health technology assessment decision-making.

We describe the key characteristics of mixture and non-mixture cure models and explain how they should be used by analysts and interpreted by reviewers and decision-makers.

## 1 Introduction

It is common for parametric survival models to be used to estimate the long-term survival benefits associated with new health technologies [1, 2]. New technologies are typically evaluated using randomised controlled trials (RCTs), but these have limited follow-up—often many participants remain alive at the end of the trial. It is therefore necessary to extrapolate to obtain complete estimates of survival benefits. This is required for health technology assessment (HTA), where lifetime costs and benefits are compared to inform healthcare resource allocation decisions [3]. Traditionally, a suite of ‘standard’ parametric survival models has been used to perform the extrapolation task [2, 4, 5], but with the development of new immuno-oncology and chimeric antigen receptor (CAR) T-cell therapies that appear to cure some patients with cancer [6–14], the use of cure models to inform HTA has attracted increasing attention [8, 9, 11, 13, 15].

There is logical rationale for the use of cure models to estimate long-term survival for treatments that may offer a cure, but these models are not commonly used to inform HTA decision-making [16]. To some extent, this lack of use may be due to a lack of understanding of how cure models work, what they assume and how reliable they are. In this tutorial we describe mixture and non-mixture cure models, focusing on key aspects that we believe analysts, reviewers

✉ Nicholas R. Latimer  
n.latimer@sheffield.ac.uk

<sup>1</sup> University of Sheffield, Sheffield, UK

<sup>2</sup> Delta Hat, Nottingham, UK

<sup>3</sup> University of Leicester, Leicester, UK

and decision-makers should be aware of when considering these models.

In Sect. 2 we provide a rationale for why cure models may be useful for HTA, as well as a brief summary of applied and methodological research on their use in an HTA context. In Sect. 3 we explain the differences between mixture and non-mixture cure models, describe how cure fractions and cure timepoints should be interpreted, explain what different cure models assume about survival in cured and uncured patients, and discuss model specifications and the use of standardised mortality ratios (SMRs) and lifetables. In Sects. 4 and 5 we provide a demonstration of the application of mixture and non-mixture cure models in a range of scenarios based on colon cancer data. In Sect. 6 we discuss scenarios in which different types of cure models may or may not be appropriate, highlighting key issues that must be taken into account by analysts when fitting these models, and by reviewers and decision-makers when interpreting their predictions. We provide code that allow the various models to be applied.

## 2 Cure Models in HTA

### 2.1 ‘Standard’ Survival Modelling and Rationale for Cure Models

For many years it has been common to use a range of ‘standard’ parametric survival models to extrapolate survival data for HTA purposes, largely because these were the focus of the first National Institute for Health and Care Excellence (NICE) Decision Support Unit (DSU) technical support document on survival analysis (TSD 14), published in 2011 [4]. However, it is well known that these models have important limitations—in particular, exponential, Weibull, Gompertz and Gamma models cannot cope with any turning points in the hazard function over time (that is, the rate at which the event of interest occurs over time), and log-logistic, log normal and Generalised Gamma models can only cope with one turning point. In some circumstances, such as when a treatment cures a proportion of patients, this can mean that these standard parametric models cannot adequately represent the hazard function.

Clinical trials usually involve strict eligibility criteria, which means that participants are unlikely to die quickly [17]. However, trials often investigate treatments for serious diseases with substantial mortality risks. Therefore, the hazard of death in an RCT may initially be low, but will rise in the short-term. Participants with the worst prognosis are likely to die first, changing the prognostic mix of those remaining in follow-up. This

may result in a turning point in the hazard function, with the hazard of death reducing in the medium term. In the long term, hazards are likely to continue to fall and may even drop to levels expected in the general population—in which case, remaining patients may be considered to be cured. In these patients, hazards will eventually rise again, due to age-related mortality risks. In this scenario, the hazard function would have two turning points, which none of the standard parametric models could accurately reflect.

This issue was the motivation for a second DSU technical support document on survival analysis, published in 2020 (TSD 21) [12]. TSD 21 describes flexible models that are able to represent complex hazard functions, including flexible parametric spline-based models, mixture models, landmark models, piecewise models, relative survival models and cure models.

### 2.2 Cure Models – the Answer?

Cure models are not commonly used to inform HTA decision-making. A review published in 2020 examined 26 NICE appraisals of immuno-oncological interventions and found that cure models were included in 8, and were considered appropriate to inform decision-making in only 3 [16]. More recently, a review of HTA reports from eight major jurisdictions for cell and gene therapies found that cure models were increasingly being used, but exact numbers were not presented [18].

The infrequent use of cure models in HTA may reflect concerns around the accuracy of their extrapolations. Simulation studies have shown that cure models can substantially over-estimate survival benefits when there is in fact not a cure, or when trial follow-up is very short [12, 15, 16]. Othus et al. analysed 6 leukaemia trials, fitting mixture cure models to initial data-cuts used in regulatory submissions and then fitting the same models to subsequent data-cuts with 3–10 years more follow-up [19]. In 4 of the 6 trials, estimates of extrapolated survival were higher based on models fitted to the original, short-term data-cut.

However, studies have also shown the potential value of cure models. Bullement et al. revisited a NICE appraisal of ipilimumab for advanced melanoma, comparing models fitted to the 3-year data-cut available at the time of the appraisal with survival subsequently observed in a later data-cut [9]. Cure models provided more accurate extrapolations than non-cure models. Similarly, Vadgama et al. compared survival predictions made by models fitted to 1-year data from the ZUMA-1 trial of axicabtagene ciloleucel for diffuse large B-cell lymphoma with survival observed in the 4-year data-cut [20]. Cure models provided accurate extrapolations,

but standard and spline-based parametric models did not. Other similar studies exist [8, 10, 21], and simulation studies have shown that cure models can extrapolate well when a cure exists and follow-up is relatively mature [15, 16].

### 2.3 Cure Models – Unanswered Questions

Two tutorial papers on the use of cure modelling in the context of HTA have been published [11, 13]. These focus on a subset of cure models – mixture cure models (MCMs). TSD 21 highlights that there are two types of cure model – MCMs and non-mixture cure models (NMCs) [12]. Each have advantages and disadvantages which may make them more or less appropriate for extrapolating from the relatively short-term, relatively small clinical trial datasets typically seen in HTA. In this tutorial, we consider both types of cure model, with the aim of providing information on how these models can be used in different situations, and how they should be interpreted.

## 3 Methods

In the context of evaluating overall survival, which is the focus of this tutorial, all cure models estimate the cure fraction (the proportion of patients who will not die from their disease), and survival for uncured patients. If cure models are used to analyse different endpoints, definitions change. For instance, if used to analyse progression-free survival, cure models estimate the proportion of patients who will not die or experience disease progression, and progression-free survival for uncured patients. Irrespective of the endpoint being analysed, MCMs and NMCs estimate cure and survival in different ways, and can be applied using different frameworks, with implications for the definition of the cure fraction. In this Section we describe frameworks for fitting cure models, key characteristics of MCMs and NMCs, model specification and the use of lifetables and SMRs.

### 3.1 Frameworks for Cure Models

It is important to determine what is meant when a cure model is used, and this depends on the framework in which the model is fitted. In the context of healthcare interventions that prevent people from dying from the disease the treatment is for, it is logical to consider cure as occurring when the all-cause hazard function for death,  $h_i(t)$ , converges with the general population hazard function,  $h_i^*(t)$ . In this context, it is typical to fit cure models in an ‘excess mortality’ framework, also known as a ‘relative survival’ framework; where we model the difference between the hazard function observed in the trial and the general population hazard (i.e. the excess hazard

that is associated with the disease of interest,  $\lambda_i(t)$ ). That is, we partition the total all-cause hazard as follows:

$$h_i(t) = h_i^*(t) + \lambda_i(t).$$

We can transform from the hazard to the survival scale, and rearrange the above equation to show that relative survival [ $R_i(t)$ ] is the ratio of the all-cause survival  $S_i(t)$  and the expected survival in the background population  $S_i^*(t)$ :

$$R_i(t) = \frac{S_i(t)}{S_i^*(t)}.$$

Using this framework, cure models are fitted to the relative survival function, which will plateau if and when the all-cause hazard approaches general population mortality rates, and the excess hazard function approaches 0. All-cause survival estimates are derived by multiplying the estimated relative survival function by the expected survival function for the background population. Taking this approach means that we directly use general population mortality rates when fitting models to the trial data, and allow these rates to govern the long-term survival function. If we expect that the disease will have a lasting effect on survival, even in cured patients, SMRs can be applied to general population mortality rates (see Sect. 3.5), and the relative survival function will plateau if and when the all-cause hazard approaches general population mortality rates with the SMR applied.

It is possible to instead fit cure models using different frameworks. A ‘disease-specific’ framework could be used, where models are fitted to disease-specific survival functions (using cause of death information), or an ‘all-cause’ framework could be used, where models are fitted to all-cause survival functions. Under either approach, estimates from these models would need to be combined with general population mortality rates in the long term. For models fit to disease-specific survival functions, this could be done using data on other-cause deaths for the trial period, and then from lifetables beyond the trial period. For models fit to all-cause survival functions it is more complex – in particular, decisions have to be made about when to build in background mortality rates: if this is done from time 0, this would double count early events; if it is done at a later timepoint questions would be asked about how the timepoint was chosen.

Because the relative survival approach is intuitive, does not need information on cause of death and avoids the need to make assumptions around when to begin incorporating general population mortality rates, we focus the rest of this tutorial on cure models fitted in a relative survival framework. This has implications for the definition of the cure fraction: in a relative survival framework, because we directly model the relative survival function, the cure

fraction corresponds to the proportion of patients alive in a world where patients can only die from the disease of interest. Therefore, while the cure fraction does represent the proportion estimated not to die from their disease, it will always be higher than the all-cause survival function at the cure timepoint (the point at which excess disease-related hazards fall to zero), because some people will have died of other causes before this timepoint.

### 3.2 Mixture Cure Models (MCMs)

MCMs assume that there are two groups of individuals – those who are cured of their disease and those who are not [22–24]. When fitted in a relative survival framework, general population mortality rates are incorporated directly into the model and the model uses these, combined with the parametric distribution chosen to represent the uncured patients, to estimate the cure fraction. General population mortality rates are taken from relevant lifetables, with rates from the appropriate calendar year used, and these are further stratified by characteristics such as age and sex, so that each trial participant can be assigned an expected background mortality rate. MCMs can be fitted using standard software packages, such as `strsmix` in Stata [25], and `flexsurv` and `cuRe` in R [26, 27].

#### 3.2.1 When Does Cure Occur?

Strictly speaking, MCMs assume that at the study baseline there is a group of patients who experience no excess mortality compared with the general population – that is, ‘cured’ patients are ‘cured’ at baseline [25]. This makes interpretation awkward, but it is also useful to consider MCMs with respect to the timepoint at which no uncured patients remain – representing the timepoint after which all remaining patients are assumed to be cured. MCMs place no constraint on this timepoint – it could occur early in the trial if the difference between hazards observed in the trial and hazards in the general population disappears quickly, or it could occur much later or not at all: sometimes MCMs will predict a 0% cure fraction. It is also important to note that the assignment of cure in an MCM is probabilistic rather than deterministic – individuals are not segregated into cured and uncured groups, they are assigned a probability of being cured and the cure fraction is estimated at the population level.

#### 3.2.2 What is Assumed for Uncured Patients?

MCMs can be fitted using a range of ‘standard’ parametric models – for example, Weibull, log normal, log-logistic, etc., to represent survival for the uncured group of patients. Hence, it is important to consider which models

are appropriate for uncured patients – for example, is it likely that the hazard function in uncured patients will be monotonically increasing or decreasing, or will have turning points? Formulations of MCMs using flexible parametric models have been developed, but are seldom used in practice [28].

#### 3.2.3 What About the Cure Fraction?

The choice of parametric distribution used within an MCM to represent uncured patients can have an important impact on the cure fraction estimated by the model. For example, if a log-logistic MCM is used, the survival distribution for uncured patients is likely to have a decreasing hazard in the long term, and the estimated cure fraction may be low because the distribution used for the uncured group is able to represent a reducing hazard and long-term survivors. In contrast, if a Weibull MCM is used, the survival distribution for uncured patients may have an increasing hazard, and the estimated cure fraction may be high, because the distribution used for the uncured group is unable to represent long-term survivors. The two models may in fact result in similar survival curves for the cured and uncured populations combined, but these survival curves would be based on very different assumptions about survival in uncured patients, and would therefore be associated with very different cure fractions.

It should not necessarily be a concern if different MCMs give very different estimates of cure fractions: this is a function of the parametric distribution chosen for uncured patients. Instead, the focus should be on selecting appropriate distributions for the uncured patient group.

### 3.3 Non-Mixture Cure Models (NMCs)

The key difference between MCMs and NMCs is that NMCs do not split the population into cured and uncured groups directly, although the cure fraction and the survival of the uncured can still be estimated from these models [24, 25, 29, 30]. NMCs can be fitted using standard parametric or flexible parametric distributions. As for MCMs, when fitted in a relative survival framework, general population mortality rates are incorporated directly into the model and should be taken from relevant lifetable sources, using appropriate calendar years and stratifying for key characteristics such as age and sex. NMCs can be fitted using standard software packages such as `strsnmix` and `stpm2` in Stata [25, 31], and `flexsurv`, `cuRe` and `rstpm2` in R [26, 27].



### 3.3.1 When Does Cure Occur?

Unlike MCMs, NMCs do not assume that there is a group of patients who are ‘cured’ at baseline. The timepoint at which cure occurs depends on when the modelled hazards converge with those observed in the general population. When fitted using standard parametric models, there is no constraint on when this convergence will occur and typically the estimates of MCMs and NMCs using the same parametric form will be similar to one another. When NMCs are applied using flexible parametric models, the analyst can specify the point at which hazards meet background population levels – these have also been referred to as latent cure models.[30]

### 3.3.2 What is Assumed for Uncured Patients?

NMCs do not split patients into cured and uncured groups, so there is not a survival model specific to uncured patients. The parametric distribution used within the NMC must be sufficiently flexible to model the survival experience of the cohort as it approaches the cure fraction.

Fitting NMCs using flexible parametric models allows for a complex hazard function to be captured prior to the cure timepoint, and provides the analyst with an additional tool to control when the cure timepoint will occur, dictated by the placement of a ‘boundary knot’ [31]. This may be useful if external data or clinical expert opinion allows for the cure timepoint to be estimated with some level of confidence. Care must be taken with this – specifying a boundary knot at (say) 5 years may mean that hazards begin to decrease rapidly much earlier, allowing gradual convergence at 5 years. To protect against an unrealistically early sharp decrease in the hazards, it may be necessary to set the boundary knot further into the future. Research has been undertaken to investigate the estimation of cure timepoints [32, 33], and this may help inform boundary knots used within flexible parametric NMCs.

### 3.3.3 What About the Cure Fraction?

The cure fractions associated with NMCs fitted with different parametric distributions are likely to differ. As for MCMs, this is to be expected. The focus should be on selecting a distribution that is likely to be adequate for representing the hazard function prior to the cure timepoint.

## 3.4 Model Specifications

MCMs and NMCs can be fitted to each treatment group independently, or with treatment group as a covariate. In HTA, it is common to fit independent survival models

to each treatment arm, due to concerns around assuming that the treatment effect (in the form of a hazard ratio for proportional hazards models, or a time ratio for accelerated failure time models) is constant over time [2, 5]. In a MCM and NMC setting, the considerations are similar. If dependent models are used, it is possible to allow a treatment effect on the cure proportion parameter, and also on the parameters that define the survival function for the uncured [24, 25]. When choosing between dependent and independent cure models, it is important to consider the validity of assumptions around the treatment effect enforced by using dependent models, to inspect modelled estimates of hazards and survival compared with the observed data, and to consider the validity of long-term extrapolations. Independently fitted models do not enforce assumptions around the treatment effect, but the treatment effect implied by the models should still be assessed, as recommended by TSD 21 [12].

It is rare to include baseline covariates in survival models used for HTA. However, this may be more relevant for cure models because of their use of lifetable data, stratified by age and sex. Over time, if older patients are more likely to die from their cancer, the relative age mix in the remaining trial population will change, and the conditional hazard function at later timepoints will be based on younger people than would have been the case if general population mortality were the only cause of death. Therefore, if disease-related deaths are likely to be associated with age (or sex), cure models that include these as baseline variables should be superior to models that exclude them. If covariates are included in the model, it will often be necessary to obtain the marginal all-cause survival function for the cohort as a whole, especially in a HTA setting where typically marginal rather than conditional survival functions are required. This can be achieved through regression standardisation and re-incorporation of general population mortality rates. Accounting for changes in age and sex distributions over time is relatively rare in survival models used for HTA – however, this can be important and more details on standardisation are available from Lambert et al. [34].

## 3.5 Standardised Mortality Ratios (SMRs)

SMRs relate to whether cancer survivors are at a higher risk of death than the age- and sex-matched general population, separate from their disease-related mortality risk, perhaps due to co-morbidities or lasting ill-effects of intensive treatment. Various studies have reported SMRs relevant for different groups of patients with cancer, ranging from values of 6 or higher [35, 36] to values close to 1 [37].

When cure models have been used in NICE appraisals, it has been common for SMRs to be applied. Recent appraisals in ovarian cancer and gastric cancer tested SMRs ranging

between 1.4 and 1.8, though these appeared to be based primarily on assumption and clinical opinion, rather than data [38–41].

SMRs can be particularly important when the cure fraction is large. It is important to assess hazards observed during trial periods to see whether they approach background levels, but if the cure timepoint has not been reached during the study, this will not provide information on whether applying an  $SMR > 1$  would be appropriate. It may be helpful to analyse relevant registry datasets that contain long-term survivors with the same disease to investigate whether those patients exhibit mortality rates that are similar to, or higher than, background population levels.

When SMRs are applied, they should be applied to the general population mortality rates in the lifetables being used, ensuring that the adjusted rates are incorporated directly into the cure models fitted. This is done in the same way, irrespective of whether MCMs or NMCs are being used.

### 3.6 Lifetables

Cure models require the use of general population lifetables, and therefore the source of lifetable data must be chosen. Lifetables are available for different countries (and sometimes regions) and are usually split by sex, calendar year and age. In the context of fitting a cure model to data from an international clinical trial, it could be argued that it is most appropriate to use different lifetables for each patient according to their country (or region) of residence, their age and sex and the year that they entered the trial. Alternatively, it could be argued that it is more relevant to use lifetables for the country the analysis is designed for (still stratified for age, sex and calendar year).

We suggest that this choice should be dictated by the specified purpose of the analysis – is the aim purely to project survival for trial participants, or is it to project survival for a different population (i.e. the population that a decision is being made for)? In practice, as shown by TSD 21, the use of ‘incorrect’ lifetables is unlikely to have a large impact on survival predictions [12], although this might not be the case if the cure fraction is large and if alternative lifetables have large differences in life expectancy.

When cure models are used, they should ideally be fitted to patient-level data. This allows for the distribution of age, sex and calendar year to be accounted for in the model, as well as how the age and sex mix in the surviving population changes over time. If patient-level data are not available, published survival curves can be digitised to reconstruct the data using commonly used methods [42], but the reconstructed dataset will not include information on

age, sex and calendar year of recruitment for each patient. Assumptions are then required to assign values for these variables for each patient (likely based on published means and distributions), which make resulting model predictions prone to additional error, though these are likely to be small.

## 4 Demonstrating the Application of Cure Models

### 4.1 Data

To illustrate the application of a range of MCMs and NMCs, we use data on 15,564 people diagnosed with colon cancer in a North European country between 1975 and 1994, with follow-up until 1994. Variables include age, sex, diagnosis date, clinical stage at diagnosis (localised, regional, distant), survival status and time. A total of 70% died during the 20 years of follow-up. We sampled from the complete dataset to generate datasets more similar to those collected in RCTs, representing a ‘medium cure’ scenario (Scenario 1), a ‘low cure’ scenario (Scenario 2) and a ‘no cure’ scenario (Scenario 3):

- (i) Scenario 1: 440 patients randomly selected from those with regional disease. Approximately 28% of these patients were alive at 10 years.
- (ii) Scenario 2: 391 patients randomly selected from those with distant disease. Approximately 5% were alive at 10 years.
- (iii) Scenario 3: 481 patients randomly selected from those with distant disease who did not live beyond 8 years.

Summary characteristics of the datasets generated for each of these scenarios are presented in the supplementary materials.

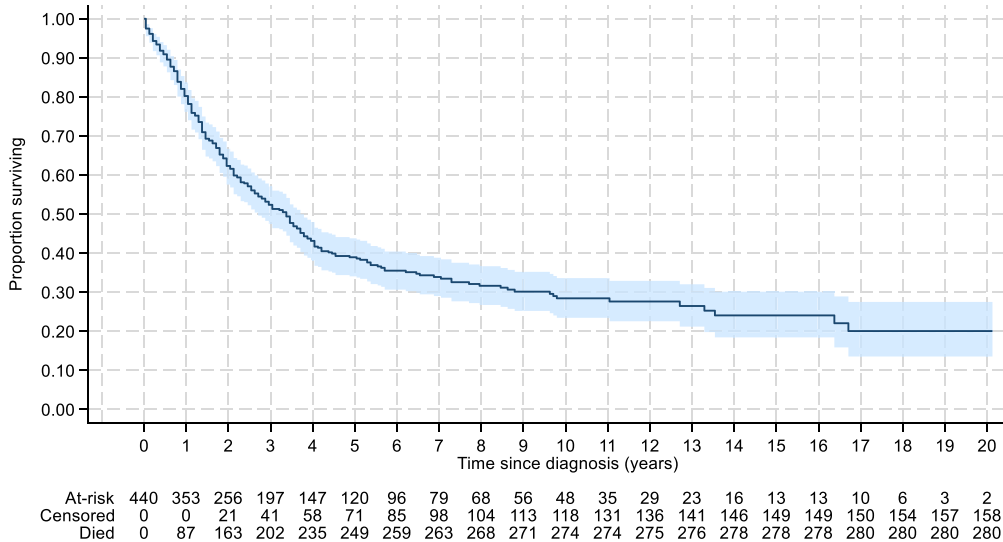
To replicate circumstances whereby models are used to extrapolate from immature RCT datasets, we conducted analyses with data artificially censored at 24 months and 48 months for each scenario. Kaplan–Meier plots for Scenarios 1–3 are presented in Fig. 1.

Dataset files and lifetable data are provided in supplementary materials. These are also available from the `strs` package in Stata [43], and the `cuRe` package in R [26].

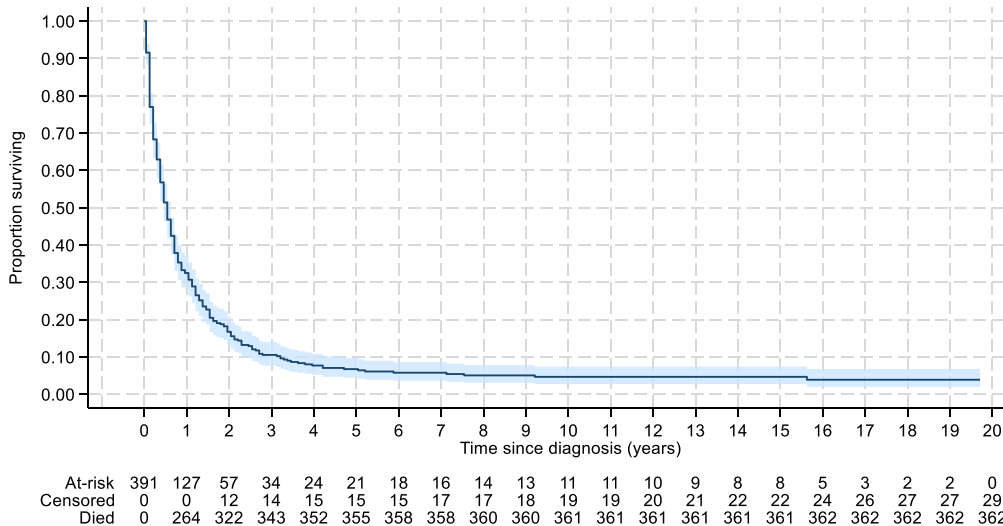
### 4.2 Models Applied

We fitted the following models to the 24- and 48-month data for each scenario:

(a) Scenario 1 (medium cure fraction)



(b) Scenario 2 (low cure fraction)



(c) Scenario 3 (no cure)

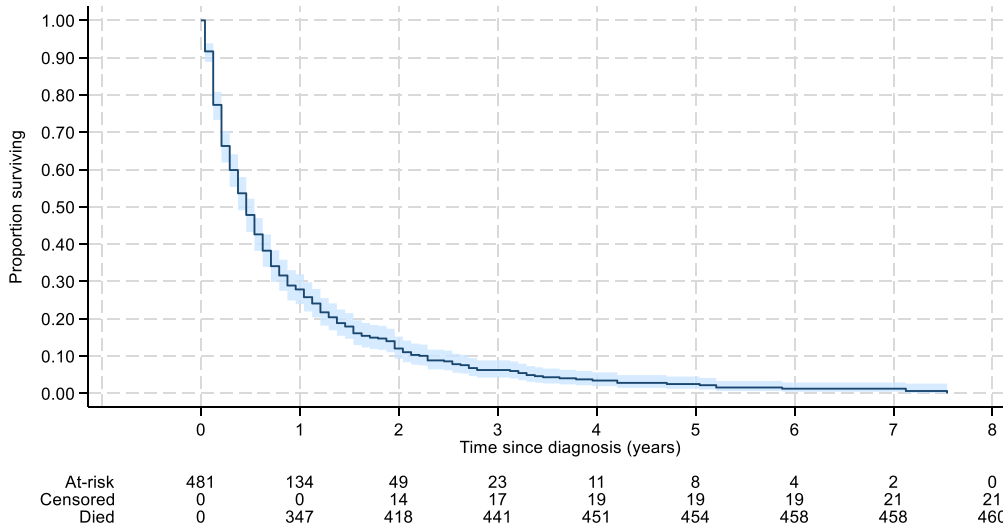


Fig. 1 Kaplan-Meier survival plots



- MCMs
  - o Log normal distribution
  - o Weibull distribution
  - o With and without including age as a baseline covariate
- Flexible parametric NMCs
  - o With a 5-year boundary knot
  - o With a 15-year boundary knot
  - o With 2 and 4 interior knots placed at default centiles of the distribution of the uncensored log survival times (50% and 95% for 2 knots; 25%, 50%, 75% and 95% for 4 knots)
  - o With and without including age as a baseline covariate

For models that included age as a baseline covariate we used standardisation to obtain marginal all-cause survival curves. Appropriate lifetable data were used to incorporate background hazards, matched by age, sex and calendar year. We conducted analyses with no adjustment to the background hazard rates (i.e. SMR = 1) and repeated this with an SMR of 2.5, where we multiplied the lifetable hazards used in the models by 2.5.

We recorded model estimates of restricted mean survival time (RMST) at 10 and 20 years, survival proportions at 5, 10 and 20 years and constructed plots of estimated hazard and survival functions. We compared these estimates with observed outcomes (without artificial censoring), and assessed whether model estimates lay within 95% confidence intervals (CI) of observed values. However, the aim of our analyses was not to provide a definitive evaluation of the performance of the models, rather, it was simply to demonstrate their application.

In the supplementary materials we provide Stata code for conducting the analyses. We fitted MCMs using `strsmix` combined with `stexpect3` [25, 44], and fitted NMCs using `stpm2` combined with `standsurv` [45, 46].

## 5 Results

Results for Scenarios 1–3 are presented in Tables 1, 2, 3 and Figs. 2, 3, 4. Results for models that used an SMR of 2.5 are presented in the supplementary materials. For ease of interpretation, in Figs. 2, 3, 4 we do not include all models. In general, including age as a baseline covariate and the number of internal knots included in NMCs made little difference to model estimates, so we include plots only for models that included age, and for NMCs with 2 internal knots. The only exception to this is for Scenario 3, where

MCMs with age as a baseline covariate did not converge: MCMs that excluded age are presented instead.

For each scenario we briefly describe the results, and then explain why the models produced these results.

### 5.1 Scenario 1 (Medium Cure Fraction)

#### 5.1.1 Model Estimates

When models were fit to 24-month data, only NMCs with a 15-year boundary knot produced estimates that fell within the 95% CIs of the observed survival proportions at 5, 10 and 20 years (Table 1, Fig. 2). MCMs substantially underestimated long-term survival, and NMCs with a 5-year boundary knot predicted survival curves that flattened too early. Including age as a baseline covariate in the models made little difference to the NMCs, but improved MCM estimates.

Estimated cure fractions ranged considerably. As explained in Sections 3.2 and 3.3, cure fractions should be interpreted with extreme care, and it is likely to be more appropriate to consider long-term all-cause survival predictions. We report both in Table 1 to illustrate this point. Cure fractions ranged from 0–21% with MCMs to 48–62% with NMCs, whilst survival at 10 years ranged from 6–21% with MCMs to 30–39% with NMCs. Observed survival at 10 years was 28% (95% CI 23–34%). RMST estimates at 20 years ranged from 3.7–5.7 years for MCMs, and from 7.1–8.3 years for NMCs, whilst the observed value was 7.1 years (95% CI 6.3–7.9).

When models were fit to the 48-month data, only NMCs with a 5-year boundary knot, and those with a 15-year boundary knot which included age in the model, produced estimates that fell within the 95% CIs of the observed survival proportions at 5, 10 and 20 years. MCMs again substantially underestimated long-term survival, although their estimates were closer to the observed values than those associated with the 24-month models. Cure fractions ranged from 0–38% with MCMs to 42–53% with NMCs, and survival at 10 years ranged from 18–23% with MCMs to 26–33% with NMCs [observed value: 28% (95% CI 23–34%)]. RMST estimates at 20 years ranged from 5.4–6.1 years for MCMs, to 6.5–7.4 years with NMCs [observed value 7.1 years (95% CI 6.3–7.9)].

#### 5.1.2 Explaining the Results

Figure 2b shows that observed hazards increased until just after 2 years, before beginning to decrease. MCMs fitted to the 2-year data could not identify the turning point in the hazard and estimated low or zero cure fractions, underestimating survival as a result. In contrast, the NMCs were ‘told’ that hazards will return to background population

**Table 1** Scenario 1 (medium cure fraction) model predictions and observed survival, SMR = 1

Model	Restricted mean survival time		Survival proportions			Cure fraction	AIC
	10 years	20 years	5 years	10 years	20 years		
Observed data (95% CI)	4.71 (4.33–5.09)	7.10 (6.27–7.93)	38.9% (34.1–43.8%)	28.4% (23.4–33.6%)	20.1% (13.5–27.5%)	N/A	N/A
<b>Models fitted to 24 months of data</b>							
MCM log normal	<b>4.35</b>	5.38	<b>37.5%</b>	18.3%	5.1%	0%	743
MCM Weibull	3.54	3.70	27.0%	5.7%	0.2%	0%	736
NMC, 5y bk, 2 knot	5.42	8.18	49.6%	37.6%	<b>18.9%</b>	62.3%	737
NMC, 5y bk, 4 knot	5.41	8.15	49.4%	37.4%	<b>18.9%</b>	62.0%	740
NMC, 15y bk, 2 knot	<b>4.88</b>	<b>7.06</b>	<b>42.0%</b>	<b>29.8%</b>	<b>15.0%</b>	49.3%	736
NMC, 15y bk, 3 knot	<b>4.88</b>	<b>7.06</b>	<b>42.0%</b>	<b>29.9%</b>	<b>15.0%</b>	49.3%	738
Models with age included as a baseline covariate							
MCM log normal	<b>4.38</b>	5.70	<b>38.3%</b>	20.5%	8.6%	14.7%	753
MCM Weibull	3.87	4.75	31.0%	14.0%	5.7%	21.3%	747
NMC, 5y bk, 2 knot	5.45	8.34	49.9%	38.7%	<b>20.4%</b>	61.6%	744
NMC, 5y bk, 4 knot	5.42	8.29	49.5%	38.3%	<b>20.2%</b>	61.0%	747
NMC, 15y bk, 2 knot	<b>4.91</b>	<b>7.23</b>	<b>42.5%</b>	<b>31.0%</b>	<b>16.3%</b>	49.0%	743
NMC, 15y bk, 3 knot	<b>4.88</b>	<b>7.15</b>	<b>41.9%</b>	<b>30.5%</b>	<b>16.1%</b>	48.2%	745
<b>Models fitted to 48 months of data</b>							
MCM log normal	<b>4.34</b>	5.36	<b>37.4%</b>	18.2%	5.0%	0%	1094
MCM Weibull	<b>4.44</b>	6.08	<b>37.0%</b>	22.9%	11.0%	37.7%	1089
NMC, 5y bk, 2 knot	<b>4.94</b>	<b>7.29</b>	<b>42.3%</b>	<b>32.0%</b>	<b>16.2%</b>	53.1%	1099
NMC, 5y bk, 4 knot	<b>4.90</b>	<b>7.21</b>	<b>41.4%</b>	<b>31.4%</b>	<b>15.8%</b>	52.0%	1098
NMC, 15y bk, 2 knot	<b>4.63</b>	<b>6.53</b>	<b>38.3%</b>	<b>26.1%</b>	13.1%	43.0%	1090
NMC, 15y bk, 4 knot	<b>4.62</b>	<b>6.50</b>	<b>38.2%</b>	<b>25.8%</b>	12.9%	42.4%	1092
Models with age included as a baseline covariate							
MCM log normal	<b>4.39</b>	5.87	<b>38.0%</b>	21.5%	10.4%	15.9%	1092
MCM Weibull	4.26	5.69	<b>36.0%</b>	19.7%	11.0%	24.4%	1090
NMC, 5y bk, 2 knot	<b>4.94</b>	<b>7.38</b>	<b>42.3%</b>	<b>32.6%</b>	<b>17.3%</b>	52.4%	1104
NMC, 5y bk, 4 knot	<b>4.90</b>	<b>7.29</b>	<b>41.4%</b>	<b>32.0%</b>	<b>17.0%</b>	51.2%	1103
NMC, 15y bk, 2 knot	<b>4.65</b>	<b>6.68</b>	<b>38.6%</b>	<b>27.1%</b>	<b>14.4%</b>	42.9%	1088
NMC, 15y bk, 4 knot	<b>4.63</b>	<b>6.61</b>	<b>38.3%</b>	<b>26.5%</b>	<b>14.1%</b>	41.8%	1091

Bold – estimate lies within 95% confidence interval of observed data. Some NMCs fitted to 24-month data that included 4 internal knots did not converge, so 3 knots were used instead

AIC Akaike information criterion, CI confidence interval, MCM mixture cure model, NMC non-mixture cure model, SMR standardised mortality ratio, 5y bk 5-year boundary knot, 15y bk 15-year boundary knot

levels at the boundary knot timepoint, and were therefore forced to predict a turning point in the hazard.

Estimates from the MCMs improved when they were fit to the 48-month data because the turning point in the hazard had occurred by this point. However, the MCMs estimated hazards that decreased too slowly, and long-term survival continued to be under-estimated.

NMCs with a 5-year boundary knot forced survival curves to flatten too quickly, because, in fact, observed hazards reached background levels at 8 years. NMCs with a 15-year boundary knot more closely approximated the decrease in hazards, but slightly underestimated long-term survival. Figure 2d illustrates that observed hazards fell below background hazards between 8 and 15 years, likely due to chance and small sample

**Table 2** Scenario 2 (low cure fraction) model predictions and observed survival, SMR = 1

Model	Restricted mean survival time		Survival proportions			Cure fraction	AIC
	10 years	20 years	5 years	10 years	20 years		
Observed data (95% CI)	1.35 (1.11–1.58)	1.77 (1.34–2.20)	6.7% (4.4–9.7%)	4.7% (2.7–7.4%)	3.9% (2.0–6.8%)	N/A	N/A
<b>Models fitted to 24 months of data</b>							
MCM log normal	<b>1.24</b>	<b>1.43</b>	<b>6.2%</b>	<b>3.0%</b>	1.2%	4.1%	534
MCM Weibull	1.63	2.23	11.5%	8.4%	<b>4.0%</b>	14.9%	554
NMC, 5y bk, 2 knot	1.81	2.53	13.5%	10.0%	<b>4.9%</b>	17.6%	550
NMC, 5y bk, 4 knot	1.72	2.38	12.2%	9.0%	<b>4.5%</b>	15.9%	540
NMC, 15y bk, 2 knot	<b>1.56</b>	<b>2.05</b>	9.9%	<b>6.8%</b>	<b>3.3%</b>	11.9%	543
NMC, 15y bk, 4 knot	<b>1.49</b>	<b>1.92</b>	<b>8.9%</b>	<b>5.9%</b>	<b>2.9%</b>	10.4%	537
Models with age included as a baseline covariate							
MCM log normal	<b>1.14</b>	1.25	<b>5.5%</b>	1.8%	0.7%	3.1%	518
MCM Weibull	1.66	2.32	12.3%	9.2%	<b>4.4%</b>	16.2%	537
NMC, 5y bk, 2 knot	1.85	2.68	13.9%	10.9%	<b>5.9%</b>	17.1%	526
NMC, 5y bk, 4 knot	1.77	2.54	12.7%	10.1%	<b>5.6%</b>	15.5%	518
NMC, 15y bk, 2 knot	<b>1.57</b>	<b>2.10</b>	9.9%	<b>7.0%</b>	<b>3.7%</b>	11.2%	520
NMC, 15y bk, 4 knot	<b>1.51</b>	<b>1.99</b>	<b>9.0%</b>	<b>6.4%</b>	<b>3.4%</b>	9.9%	516
<b>Models fitted to 48 months of data</b>							
MCM log normal	<b>1.19</b>	<b>1.34</b>	<b>5.5%</b>	2.4%	0.9%	3.0%	639
MCM Weibull	<b>1.30</b>	<b>1.63</b>	<b>6.6%</b>	<b>4.7%</b>	<b>2.2%</b>	8.3%	664
NMC, 5y bk, 2 knot	<b>1.52</b>	<b>2.05</b>	9.8%	<b>7.3%</b>	<b>3.6%</b>	12.8%	672
NMC, 5y bk, 4 knot	<b>1.43</b>	<b>1.87</b>	<b>8.2%</b>	<b>6.1%</b>	<b>3.0%</b>	10.7%	653
NMC, 15y bk, 2 knot	<b>1.37</b>	<b>1.73</b>	<b>7.5%</b>	<b>4.9%</b>	<b>2.4%</b>	8.6%	650
NMC, 15y bk, 4 knot	<b>1.31</b>	<b>1.60</b>	<b>6.4%</b>	<b>4.1%</b>	<b>2.0%</b>	7.1%	642
Models with age included as a baseline covariate							
MCM log normal	<b>1.12</b>	1.23	<b>5.2%</b>	1.9%	0.7%	2.1%	624
MCM Weibull	<b>1.34</b>	<b>1.73</b>	<b>7.7%</b>	<b>5.3%</b>	<b>2.6%</b>	10.6%	651
NMC, 5y bk, 2 knot	<b>1.55</b>	<b>2.16</b>	10.1%	7.9%	<b>4.3%</b>	12.4%	648
NMC, 5y bk, 4 knot	<b>1.47</b>	<b>1.99</b>	<b>8.6%</b>	<b>6.9%</b>	<b>3.8%</b>	10.4%	633
NMC, 15y bk, 2 knot	<b>1.38</b>	<b>1.77</b>	<b>7.6%</b>	<b>5.2%</b>	<b>2.7%</b>	8.2%	628
NMC, 15y bk, 4 knot	<b>1.34</b>	<b>1.68</b>	<b>6.7%</b>	<b>4.5%</b>	<b>2.4%</b>	6.9%	621

Bold – estimate lies within 95% confidence interval of observed data

AIC Akaike information criterion, CI confidence interval, MCM mixture cure model, NMC non-mixture cure model, SMR standardised mortality ratio, 5y bk 5-year boundary knot, 15y bk 15-year boundary knot

sizes. Therefore, it is reasonable to conclude that NMCs with 15-year boundary knots provided good predictions when fitted to 24- and 48-month data in this scenario. Notably, cure fractions estimated by these models were substantially higher than the proportion alive at the 10-year timepoint because, as explained in Sect. 3.1, the cure fraction corresponds to the relative survival function, not the all-cause survival function.

Models that included an SMR of 2.5 substantially underestimated long-term survival (Table S2, Fig. S1) because observed hazards fell to background levels: an SMR > 1 was not appropriate.

## 5.2 Scenario 2 (Low Cure Fraction)

### 5.2.1 Model Estimates

When fitted to 24-month data, all models except the log normal MCM appeared to estimate too sharp a decrease in the hazard function between 1 and 6 years (Fig. 3b), resulting in over-estimated survival (Table 2, Fig. 3). The log normal MCM provided good estimates of survival at 5 years, but under-estimated survival in the longer term. Cure fractions ranged from 3–16% with MCMs to 10–18% with NMCs, and survival at 10 years ranged from 2% with the log normal MCM to 9% with the Weibull MCM, and from 6 to 11% with NMCs, compared with the observed value of 5% (95% CI 3–7%). RMST estimates at 20 years ranged from 1.3 with

**Table 3** Scenario 3 (no cure) model predictions and observed survival, SMR = 1

Model	Restricted mean survival time		Survival proportions			Cure fraction	AIC
	10 years	20 years	5 years	10 years	20 years		
Observed data (95% CI)	0.91 (0.80–1.02)	0.91 (0.80–1.02)	2.5% (1.2–4.5%)	0%	0%	N/A	N/A
<b>Models fitted to 24 months of data</b>							
MCM log normal	<b>0.90</b>	<b>0.92</b>	<b>2.6%</b>	<b>0.6%</b>	<b>0.1%</b>	0.3%	582
MCM Weibull	1.27	1.68	7.8%	5.7%	2.7%	10.3%	609
NMC, 5y bk, 2 knot	1.46	1.99	9.9%	7.3%	3.5%	13.1%	604
NMC, 5y bk, 4 knot	1.37	1.82	8.6%	6.3%	3.0%	11.3%	593
NMC, 15y bk, 2 knot	1.24	1.55	6.7%	4.4%	2.1%	7.9%	593
NMC, 15y bk, 4 knot	1.16	1.42	5.6%	3.6%	1.7%	6.4%	588
Models with age included as a baseline covariate							
MCM log normal	.	.	.	.	.	.	.
MCM Weibull	.	.	.	.	.	.	.
NMC, 5y bk, 2 knot	1.49	2.08	10.2%	7.8%	4.2%	12.8%	580
NMC, 5y bk, 4 knot	1.40	1.93	8.9%	6.9%	3.8%	11.1%	571
NMC, 15y bk, 2 knot	1.24	1.57	6.7%	4.5%	2.3%	7.6%	569
NMC, 15y bk, 4 knot	1.17	1.45	5.7%	3.8%	2.0%	6.2%	566
<b>Models fitted to 48 months of data</b>							
MCM log normal	<b>0.87</b>	<b>0.89</b>	<b>2.3%</b>	<b>0.4%</b>	<b>0.0%</b>	0.0%	672
MCM Weibull	<b>0.94</b>	1.08	<b>2.9%</b>	2.0%	0.9%	3.6%	708
NMC, 5y bk, 2 knot	1.21	1.57	6.8%	5.0%	2.4%	8.9%	725
NMC, 5y bk, 4 knot	1.11	1.39	5.2%	3.8%	1.8%	6.8%	702
NMC, 15y bk, 2 knot	1.06	1.26	<b>4.5%</b>	2.8%	1.4%	5.0%	693
NMC, 15y bk, 4 knot	<b>1.00</b>	1.14	<b>3.4%</b>	2.0%	1.0%	3.5%	681
Models with age included as a baseline covariate							
MCM log normal	<b>0.87</b>	<b>0.94</b>	<b>3.1%</b>	1.1%	0.5%	1.8%	655
MCM Weibull	<b>0.92</b>	1.06	<b>3.3%</b>	2.1%	0.9%	4.5%	694
NMC, 5y bk, 2 knot	1.23	1.62	6.9%	5.3%	2.8%	8.7%	700
NMC, 5y bk, 4 knot	1.14	1.45	5.4%	4.2%	2.3%	6.7%	681
NMC, 15y bk, 2 knot	1.06	1.27	<b>4.5%</b>	2.8%	1.4%	4.9%	667
NMC, 15y bk, 4 knot	<b>1.01</b>	1.17	<b>3.5%</b>	2.1%	1.1%	3.6%	658

Bold – estimate lies within 95% confidence interval of observed data. MCMs fitted to 24-month data that included age as a baseline covariate did not converge.

AIC Akaike information criterion, CI confidence interval, MCM mixture cure model, NMC non-mixture cure model, SMR standardised mortality ratio, 5y bk 5-year boundary knot, 15y bk 15-year boundary knot

the log normal MCM to 2.3 years for the Weibull MCM, and from 1.9 to 2.7 years with NMCs, whilst the observed value was 1.8 years (95% CI 1.3–2.2). Including age as a baseline covariate made little difference.

When fitted to the 48-month data, models performed much better: in particular, NMCs with a 15-year boundary knot and Weibull MCMs produced estimates that closely approximated observed survival proportions at 5, 10 and 20 years. Log normal MCMs continued to under-estimate long-term survival, and NMCs with a 5-year boundary knot tended to over-estimate survival. Cure fractions ranged from 2–10% with MCMs to 7–13% with NMCs, and survival at 10 years ranged from 2% with the log normal MCM and 5% with the Weibull MCM, and from 4

to 8% with NMCs [observed value: 5% (95% CI 3–7%)]. RMST estimates at 20 years ranged from 1.2 with the log normal MCM to 1.7 years for the Weibull MCM, and from 1.6 to 2.2 years with NMCs [observed value 1.8 years (95% CI 1.3–2.2)].

### 5.2.2 Explaining the Results

Figure 3b illustrates that the turning point in the observed hazard occurred more quickly in Scenario 2 than in Scenario 1. MCMs were able to identify the turning point and estimate reducing long-term hazards, similar to those predicted by the NMCs, even when fitted to the 24-month data. Hence, the range in estimates from the different cure models was

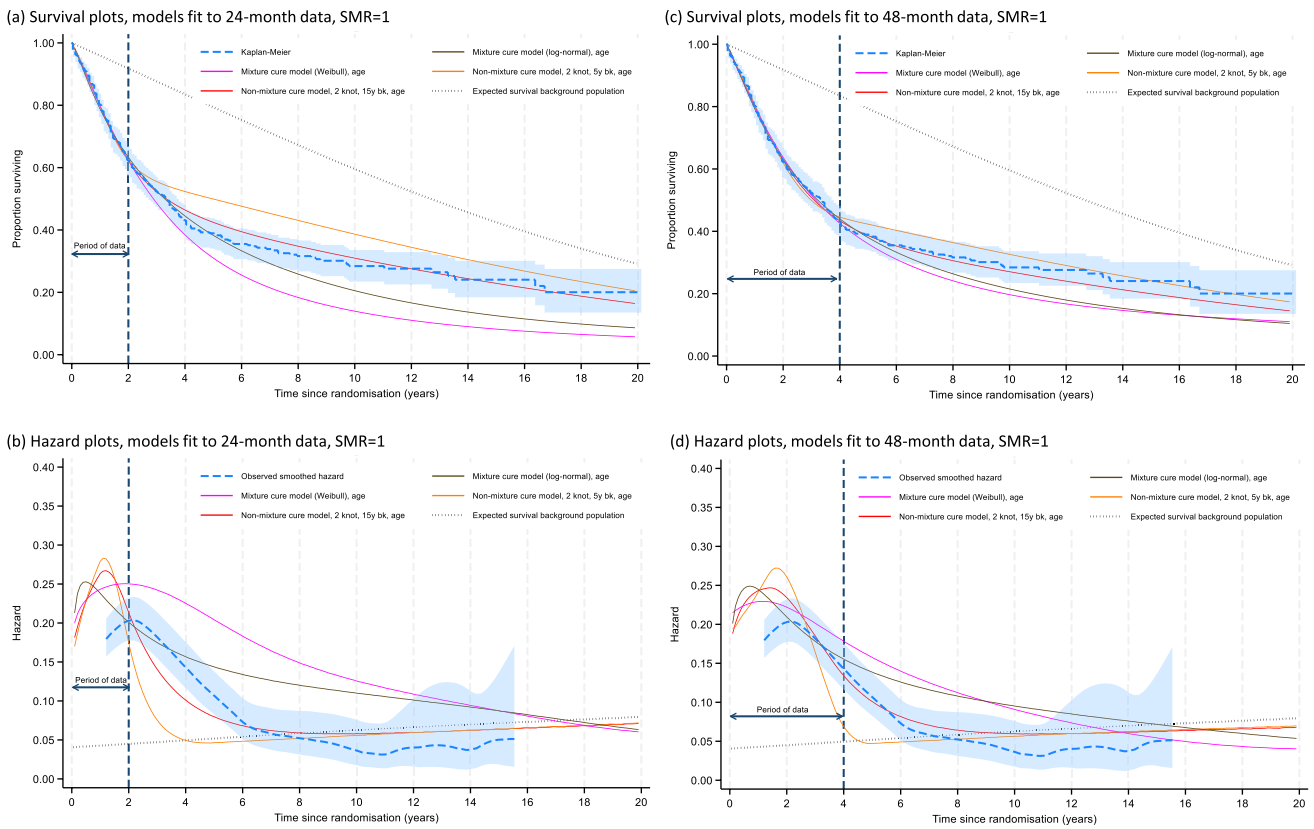


Fig. 2 Scenario 1 (medium cure fraction) survival and hazard plots. 5y bk 5-year boundary knot, 15y bk 15-year boundary knot

narrower in Scenario 2. However, none of the models fitted to the 24-month data could accurately predict the long-term gradient of the decrease in the hazards. This was vastly improved when models were fit to the 48-month data, when the gradient of the decrease was more established. NMCs with a 5-year boundary knot forced hazards to decline too quickly, whereas those with a 15-year boundary knot predicted the observed hazards closely. These and the Weibull MCM produced credible estimates of long-term survival in this scenario.

### 5.3 Scenario 3 (No Cure)

In this scenario all models except the log normal MCM substantially over-estimated survival at 5, 10 and 20 years when fitted to the 24-month data, with estimates slightly improved when using the 48-month data. The log normal MCM produced accurate survival estimates because it predicted a low cure fraction and its distribution provided a reasonable approximation of the observed hazards.

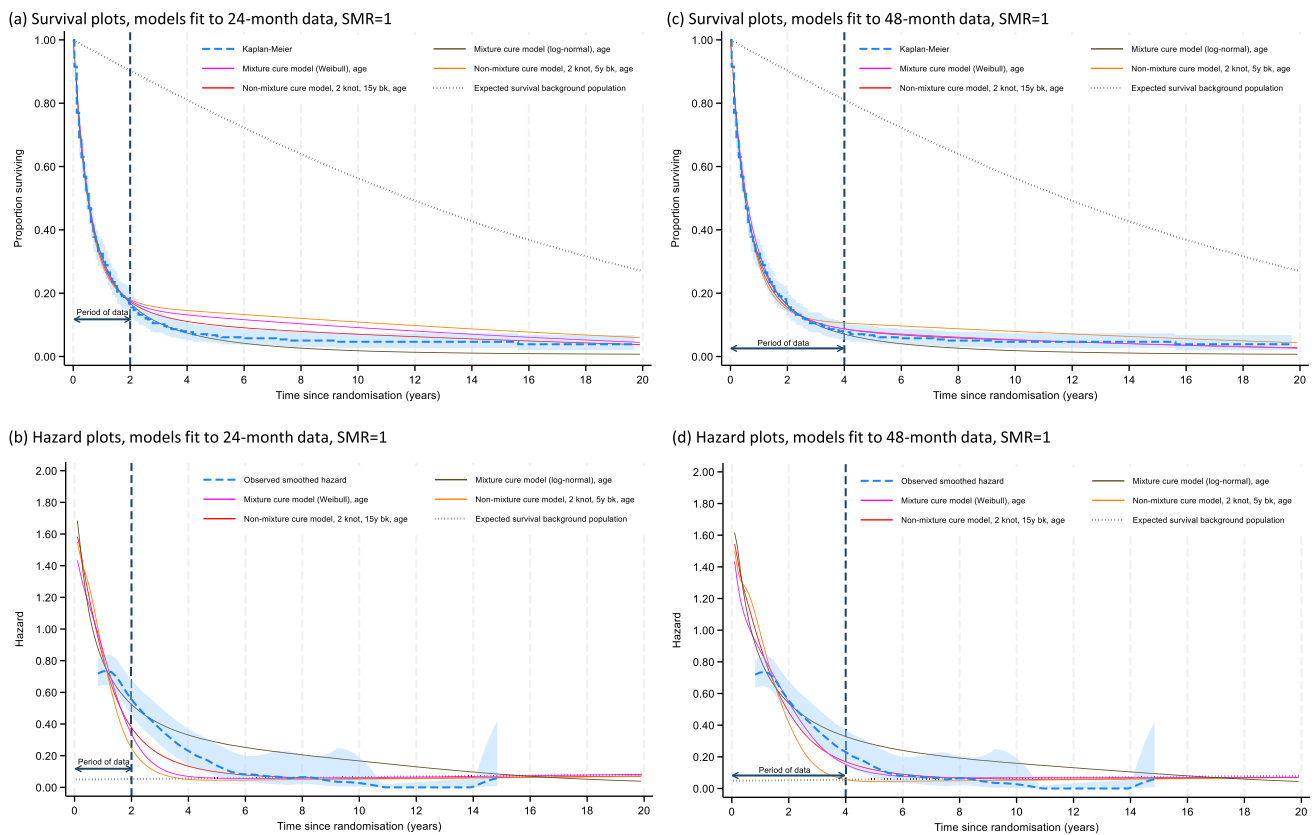
## 6 Discussion

In this tutorial we have outlined key characteristics of MCMs and NMCs, and demonstrated their application in three scenarios. Here we outline aspects to consider when using these models.

### 6.1 Think About the Hazard Function

If hazards in the observed period are increasing, MCMs will estimate a 0% cure fraction and extrapolations will be based on the distribution assigned to uncured patients. There is no value to fitting MCMs in this case. Flexible parametric NMCs are forced to predict a turning point in the hazard function, so that hazards return to background levels at the boundary knot timepoint. Therefore, if hazards are increasing in the observed period, but a cure can be confidently predicted (perhaps on the basis of external data, or strong clinical opinion), flexible parametric NMCs are the only logical cure model option.

When hazards have begun decreasing during follow-up, MCMs and NMCs may produce credible extrapolations, but this will depend on how well established the decreasing hazard is, and it is difficult to determine whether data are mature enough for the shape of the hazard function



**Fig. 3** Scenario 2 (low cure fraction) survival and hazard plots. *5y bk* 5-year boundary knot, *15y bk* 15-year boundary knot

to be well enough established for models to extrapolate accurately. With less severe disease the rate of events will be lower and it may take longer for the shape of the hazard function to become established: in general, longer-term follow-up is required for less severe diseases, if cure models are to be relied upon to provide accurate extrapolations. If hazards are observed to begin to decrease only shortly before the end of follow-up, the decrease in the hazards may not be well enough established for models to extrapolate accurately. In this case, applying flexible parametric NMCs with a range of boundary knots could be used as a form of sensitivity analysis, as the boundary knots influence the rate of decline of the extrapolated hazards.

In our analyses, a range of MCMs and NMCs extrapolated accurately in Scenario 2 when fitted to 48-month data, when 88% of patients had died and the hazard function was well established. This was not the case in Scenario 1, when only 48% had died at 48 months, the turning point in the hazard function was later, and the subsequent decrease in hazards was less well established. However, it is notable that NMCs with 15-year boundary knots *did* extrapolate well in Scenario 1, even when fitted to 24-month data. This indicates that when a cure

assumption is valid but data are heavily censored, flexible parametric NMCs may be more likely to extrapolate credibly than MCMs, though uncertainty will be high and the choice of boundary knots will be important.

In addition, when considering the hazard function and model fits, it is important to interpret plots with care. The plots of observed hazards that we present use a smoothing function because unsmoothed plots are likely to be volatile. With or without smoothing, plots of observed hazards may not provide reliable values towards the tail of the curves where numbers at risk are low. In addition, because the hazard is an instantaneous event rate rather than a cumulative measure, models may appear to provide a worse fit when comparing their predictions with observed hazards than when comparing their predictions with observed survival. Hence, a model that does not closely follow a smoothed plot of the observed hazards for its entire length does not necessarily represent a poor model, though it is desirable for model estimates to reflect any clear turning points in the observed hazards and to lie within the confidence intervals of the smoothed plots.



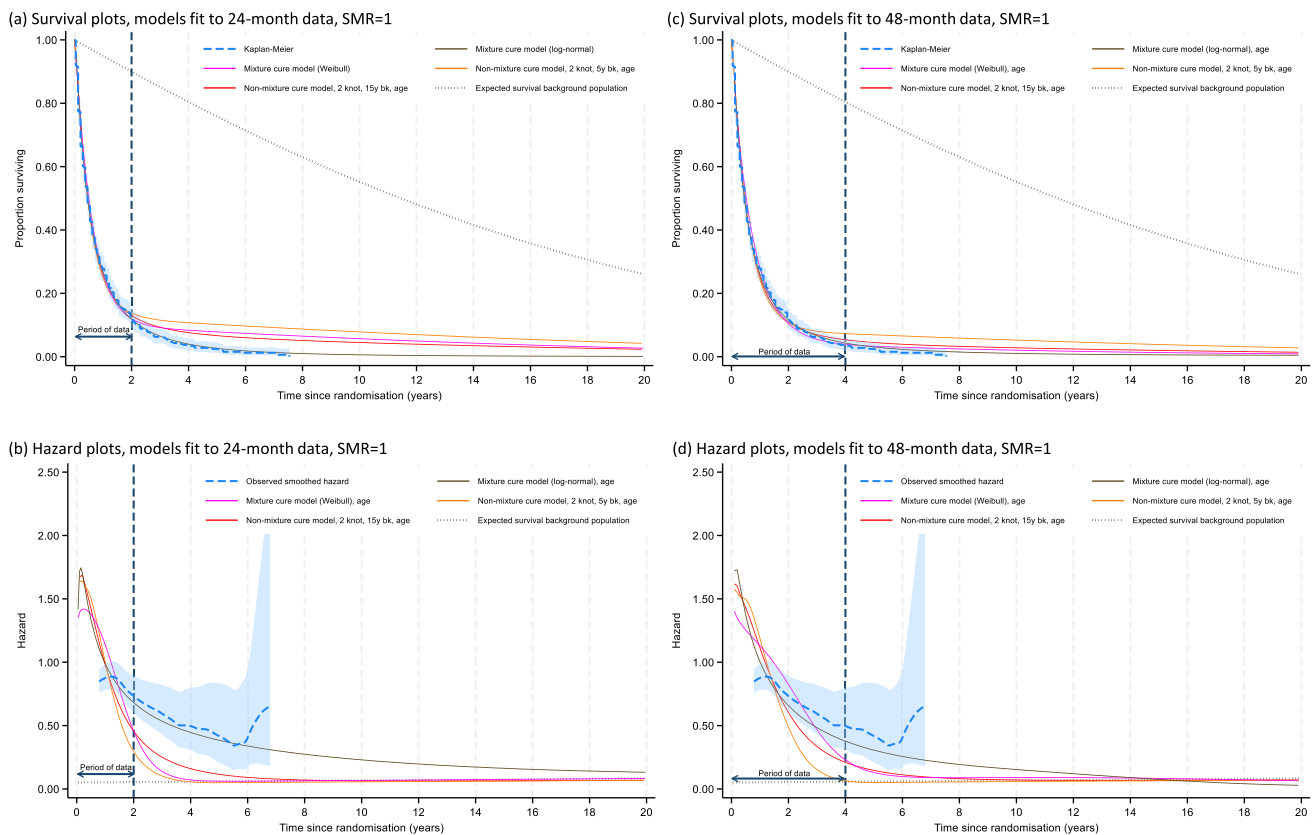


Fig. 4 Scenario 3 (no cure) survival and hazard plots. *5y bk* 5-year boundary knot, *15y bk* 15-year boundary knot

## 6.2 Choosing Boundary Knots for Flexible Parametric NMCs

Predictions from NMCs may vary substantially, depending on where the boundary knot – that is, the cure timepoint – is placed. Our analyses demonstrate that predicted hazards will begin to fall sharply substantially before the boundary knot. Setting a 5-year boundary knot resulted in hazards that fell steeply at around 2 years, so that they were almost at background levels by 3.5 years. Setting a 15-year boundary knot resulted in hazards that were not much higher than background levels at 5 years, but they fell more gradually. The shape of the hazard should be discussed with clinical experts using recognised elicitation techniques [47, 48], and when cure timepoints are uncertain, sensitivity analysis is likely to be useful. However, given that flexible parametric NMCs are likely to result in hazards that become close to background levels well before the boundary knot timepoint, setting boundary knots at relatively late timepoints is advisable.

## 6.3 Think About Survival in Uncured Patients

If MCMs are deemed appropriate, the shape of the hazard function in uncured patients should dictate the choice of MCM distribution. If hazards in uncured patients are likely to be monotonic, a Weibull MCM may be reasonable; if hazards may increase and then decrease, a log normal, log-logistic or Generalised Gamma MCM may be more appropriate. For NMCs, the chosen distribution must be able to represent the hazards before the cure timepoint. In our analyses, the number of interior knots included in the flexible parametric NMCs made little difference to model estimates. In a non-cure setting, the number of knots included in flexible parametric models dictates the portion of data upon which extrapolations are based, and therefore the number of knots chosen is often very important. In a cure setting, boundary knots have the largest influence on extrapolations – though it remains important to select a reasonable number of interior knots to capture the shape of the hazard function without over-fitting to the data [12]. In general, flexible parametric NMCs should allow a reasonable fit to observed data whilst also allowing the cure timepoint to be defined: this seems particularly useful in an HTA context.

## 6.4 Interpreting Cure Fractions

Cure fractions are a function of the framework models are fitted in, and the distributions used within the models. They should be expected to differ between models.

Cure models can predict long-term survivors even when estimating a 0% cure fraction if the distribution used for the uncured has a long tail. For cure models fitted in a relative survival framework, the cure fraction corresponds to the relative survival function and will always be greater than the all-cause survival function. The extent to which this is the case will depend on the relative importance of other cause mortality. Fundamentally, we recommend focusing on the survival proportions predicted by each model over time, rather than the estimated cure fraction.

## 6.5 SMRs and Model Specification

Applying inappropriate SMRs can lead to poor extrapolations, especially when there is a large proportion of long-term survivors. It may not be appropriate to apply a  $SMR > 1$ , as shown by our case study, especially in Scenarios 1 and 2, where observed hazards fell to general population levels. However, there is nuance to this. In Scenario 2, a proportion of patients were cured, and hazards fell to general population levels. Several of the cure models over-estimated the cure fraction, and their long-term survival predictions improved when a SMR of 2.5 was used (see supplementary materials), because this forced survival curves downwards. This is a case of two modelling errors causing bias in opposite directions and to some extent cancelling out. We believe it is preferable to attempt to determine which cure models produce credible survival extrapolations when a realistic SMR is applied, rather than applying an increased SMR in an attempt to protect against over-estimated cure fractions. SMRs should be considered on a case-by-case basis, informed by external data (ideally including an assessment of hazards in long-term survivors with the same condition) and clinical expert opinion.

In epidemiological research, it is standard practice to include age in relative survival models, including cure models [49]. This made little difference in our analyses, in which we focused on average survival curves, but it is likely to improve model accuracy within subgroups by age and is more important when there is a large proportion of long-term survivors.

## 6.6 Choosing Between Models

Survival models used for extrapolation should never be chosen purely on the basis of statistical fit to the observed data – the plausibility of extrapolations is more important [4, 5, 50]. This is especially the case for cure models, where

expectations around credible ranges of long-term survival proportions and cure timepoints are crucial. In Scenario 1, MCMs and NMCs had similar Akaike Information Criterion values but produced widely diverging extrapolations – NMCs extrapolated adequately, but MCMs did not.

## 6.7 Fitting Cure Models When There is no Cure

Fitting any type of cure model when there is in fact not a cure is likely to result in over-estimated long-term survival. MCMs with distributions for the uncured that can reflect decreasing hazards may protect against this, because these models are more likely to estimate zero cure fractions. However, in our analyses, log normal MCMs were amongst the worst performing models when there was a cure. Therefore, it is inadvisable to fit log normal MCMs ‘just in case’ there is not a cure. Instead, when cure is uncertain, it is more sensible to fit cure and non-cure models to allow reviewers and decision-makers to assess the sensitivity of effectiveness and cost-effectiveness estimates to the cure assumption.

## 6.8 Other Modelling Approaches

This tutorial focuses on MCMs and NMCs, but these are not the only models that can extrapolate cure-like long-term survival. TSD 21 describes other models, including piecewise, landmark-based and relative survival models [12]. Each approach has advantages and disadvantages and this tutorial is not an endorsement of any particular method. Landmark-based piecewise models may combine a parametric curve up until a specified landmark timepoint and general population mortality rates beyond the landmark, and in some respects may appear similar to an NMC. However, these models do not use the data to estimate the cure, and may result in discontinuities in the predicted hazard function. They also incorporate background mortality rates in a less sophisticated way. Therefore, if this type of approach is to be taken, we suggest that using a flexible parametric NMC with carefully selected boundary knots (and sensitivity analyses) would be preferable.

It is important to acknowledge that in this demonstration we present NMCs using flexible parametric models, with extra information added in the form of boundary knot timepoints, but only present MCMs using standard parametric models. This may place NMCs at an advantage. The rationale for this is that flexible parametric NMCs are well recognised for the analysis of population-level survival studies [12, 31, 51, 52], whereas, to the best of our knowledge, MCMs have very rarely been applied using flexible parametric models.

In addition, Felizzi et al. developed ‘informed’ MCMs, whereby the cure fraction is obtained from external

information and used as an input to the MCM [11]. Like using boundary knots within flexible parametric NMCs, this allows the analyst some control over the survival curve estimated by the model, and is worthy of further consideration. However, it is arguable whether it is preferable to use a cure fraction as an input to a cure model or a cure timepoint (as used within flexible parametric NMCs), especially given the difficulties associated with interpreting cure fractions estimated using different frameworks (i.e. relative survival or all-cause mortality), and the differences between fractions estimated by different parametric models. We have not included these models in this tutorial because they necessitate consideration of several additional factors and a separate tutorial focused solely on these models exists [11].

Botta et al. also introduced a potentially valuable extension to MCMs, whereby they attempted to estimate and adjust for the increased risk of non-cancer deaths in patients with cancer. This is similar in concept to using SMRs, but where the increased risk is estimated using the data to which the cure models are applied [53].

## 6.9 Limitations

In addition to limiting our focus to a specific subset of cure models that can be used to extrapolate survival, we have not applied an exhaustive set of analyses for the cure models that we have tested. We applied MCMs using a subset of standard parametric models, and NMCs using flexible parametric models. This is because we expect MCMs and NMCs using the same standard parametric models to give similar results. NMCs that use flexible parametric models provide the analyst with an additional tool with which to ‘control’ long-term survival predictions through the setting of boundary knots, and therefore we considered it valuable to demonstrate the use of these models. For simplicity, and to avoid presenting results for a very large number of analyses, we chose to apply MCMs using only Weibull and log normal distributions. These represent models where there can (the log normal MCM) and cannot (the Weibull MCM) be a turning point in the hazard function for the uncured population.

We also limited the specification of the models we fitted with respect to baseline covariates (including or excluding age) and SMRs (SMR of 1 or 2.5). We could also have included sex as a baseline covariate. Standardisation would then be required to obtain marginal survival functions, as we demonstrated for analyses that included age as a baseline covariate. In epidemiological research it is generally considered more important to include age in relative survival models, rather than sex. However, in some circumstances, such as when survival differs considerably by sex, including sex could be important.

With respect to SMRs, we did not undertake a review or an expert elicitation process to determine a potentially relevant value – in practice this should be done. We did, however, observe that in the long-term data that we derived our scenario datasets from, hazards appeared to return to background levels (albeit with associated uncertainty), indicating that a SMR of 1 is likely to be reasonable.

Finally, in the datasets constructed for each of the scenarios we analysed, for the 24- and 48-month analyses we simply censored data for patients whose observed follow-up was greater than 24 (or 48) months. Given that entry into clinical trials is usually staggered over a period of time, our approach means that there will be less uncertainty in the tails of the observed survival curves in our scenarios than there might be in a clinical trial setting. More sparse long-term data is likely to mean that hazard functions take longer to become established, which is relevant when considering the results of the analyses that we demonstrate. However, this does not negate our conclusions or interpretation, and we refer readers particularly to Sect. 6.1 for a discussion on data maturity in relation to cure modelling.

## 6.10 Further Research

As commented on in Sects. 6.8 and 6.9, this tutorial is not exhaustive with respect to models that can extrapolate cure-like long-term survival, and does not attempt to formally evaluate the performance of these methods. A neutral comparison study of the relevant model types would be valuable [54]. Demonstrating the use of flexible parametric NMCs in a range of clinical trial datasets would also be of value, as would research into appropriate methods for determining SMRs for a range of cancers. Research into cure timepoints for different cancers (and stages of cancer) would also be highly valuable – for instance, analysing timepoints at which mortality hazards reach background population levels in registry datasets with long-term follow-up.

## 7 Conclusions

Cure models are generally interpreted with extreme caution in HTA. This is reasonable because these models can produce highly variable extrapolations and are likely to be extremely inaccurate if a cure assumption is not valid. However, when a cure assumption is credible, it is reasonable to explore extrapolations from cure models. When a cure exists and data are relatively mature, with well-established hazard functions, MCMs and NMCs are likely to produce similar, credible extrapolations. However, at the time of HTA submission, survival data are usually immature. In such cases, standard parametric MCMs are unlikely to

be able to extrapolate accurately. Flexible parametric NMCs are more able to produce accurate extrapolations in cure scenarios when trial follow-up is short, provided that sensible and reasonably accurate cure timepoints are selected. Therefore, flexible parametric NMCs are likely to be more useful than standard MCMs in the context of HTA. However, extrapolations from these models will be prone to substantial uncertainty in such scenarios, and their validity rests on the credibility of the cure assumption and the placement of boundary knots.

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s40273-024-01406-7>.

**Data availability** Dataset files and lifetable data are provided in supplementary materials. These are also available from the strS package in Stata [43], and the cuRe package in R [26].

**Declarations** For the purpose of open access, the author has applied a Creative Commons Attribution (CC BY) licence to the Author Accepted Manuscript version arising from this submission.

**Funding** This work was supported by a Cancer Research UK grant for M.J.R. (grant C41379/A27583). M.J.R. was further supported by the National Institute for Health and Care Research (NIHR) Applied Research Collaboration East Midlands (ARC EM) and Leicester NIHR Biomedical Research Centre (BRC). N.R.L. is supported by Yorkshire Cancer Research (Award reference no. S406NL). The views expressed are those of the authors and not necessarily those of the NIHR, the Department of Health and Social Care or Yorkshire Cancer Research.

**Conflict of Interest** N.R.L.'s institutions have received consultancy fees and funding, unrelated to the submitted work, from BMS, Merck, Sharpe and Dohme, and from Daiichi Sankyo, GSK, Janssen, Amgen, Sanofi, for projects N.R.L. has worked on.

**Author Contributions** Both authors contributed to the study conception and design. Data collection and analysis were performed by N.R.L. with support from M.J.R. The first draft of the manuscript was written by N.R.L. and M.J.R. critically revised the work. Both authors read and approved the final manuscript.

**Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License, which permits any non-commercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc/4.0/>.

## References

- Adler AI, Latimer NR. Adjusting for nonadherence or stopping treatments in randomized clinical trials. *JAMA*. 2021;325(20):2110–1.
- Bell Gorrod H, et al. A review of survival analysis methods Used in NICE technology appraisals of cancer treatments: consistency, limitations, and areas for improvement. *Med Decis Making*. 2019;39(8):899–909.
- National Institute for Health and Care Excellence. NICE health technology evaluations: the manual. NICE; 2022.
- Latimer N. NICE DSU Technical Support Document 14: undertaking survival analysis for economic evaluation alongside clinical trials - Extrapolation with patient-level data. 2011, NICE Decision Support Unit.
- 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(6):743–54.
- Chen TT. Statistical issues and challenges in immuno-oncology. *J Immunother Cancer*. 2013;1:18.
- Melenhorst JJ, et al. Decade-long leukaemia remissions with persistence of CD4+ CAR T cells. *Nature*. 2022;602(7897):503–9.
- Ouwens M, et al. Estimating lifetime benefits associated with immuno-oncology therapies: challenges and approaches for overall survival extrapolations. *Pharmacoeconomics*. 2019;37(9):1129–38.
- Bulleman A, Latimer NR, Bell Gorrod H. Survival extrapolation in cancer immunotherapy: a validation-based case study. *Value Health*. 2019;22(3):276–83.
- Othus M, et al. Accounting for cured patients in cost-effectiveness analysis. *Value Health*. 2017;20(4):705–9.
- Felizzi F, et al. Mixture cure models in oncology: a tutorial and practical guidance. *Pharmacoeconomics - Open*. 2021;5(2):143–55.
- Rutherford MJ et al. NICE DSU Technical Support Document 21. Flexible Methods for Survival Analysis. 2020, NICE Decision Support Unit.
- Sweeting MJ, et al. Survival extrapolation incorporating general population mortality using excess hazard and cure models: a tutorial. *Med Decis Making*. 2023;43(6):737–48.
- Gye A, Goodall S, De Abreu Lourenco R. A systematic review of health technology assessments of Chimeric antigen receptor T-cell therapies in young compared with older patients. *Value Health*. 2022;25(1):47–58.
- Kearns B, et al. The extrapolation performance of survival models for data with a cure fraction: a simulation study. *Value Health*. 2021;24(11):1634–42.
- Grant TS, et al. A case study examining the usefulness of cure modelling for the prediction of survival based on data maturity. *Pharmacoeconomics*. 2020;38(4):385–95.
- Jin S, Pazdur R, Sridhara R. Re-evaluating eligibility criteria for oncology clinical trials: analysis of investigational new drug applications in 2015. *J Clin Oncol*. 2017;35(33):3745–52.
- Drummond M, et al. How are health technology assessment bodies responding to the assessment challenges posed by cell and gene therapy? *BMC Health Serv Res*. 2023;23(1):484.
- Othus M, et al. Bias in mean survival from fitting cure models with limited follow-up. *Value Health*. 2020;23(8):1034–9.
- Vadgama S, et al. Predicting survival for chimeric antigen receptor T-cell therapy: a validation of survival models using follow-up data from ZUMA-1. *Value Health*. 2022;25(6):1010–7.
- Peterse EFP, et al. Retrospective comparison of survival projections for CAR T-Cell therapies in large B-cell lymphoma. *Pharmacoeconomics - Open*. 2023;7(6):941–50.

22. De Angelis R, et al. Mixture models for cancer survival analysis: application to population-based data with covariates. *Stat Med*. 1999;18(4):441–54.
23. Boag JW. Maximum likelihood estimates of the proportion of patients cured by cancer therapy. *J Royal Stat Soc Ser B (Methodological)*. 1949;11(1):15–53.
24. Sposto R. Cure model analysis in cancer: an application to data from the Children's Cancer Group. *Stat Med*. 2002;21(2):293–312.
25. Lambert PC. Modeling of the cure fraction in survival studies. *Stata J*. 2007;7(3):351–75.
26. Jensen RK, et al. Fitting parametric cure models in R using the packages *cuRe* and *rstpm2*. *Comput Methods Programs Biomed*. 2022;226: 107125.
27. Jackson C. *flexsurv*: a platform for parametric survival modeling in R. *J Stat Software*. 2016;70(8):1–33.
28. Jakobsen LH, et al. Estimating the loss of lifetime function using flexible parametric relative survival models. *BMC Med Res Methodol*. 2019;19(1):23.
29. Lambert PC, et al. Estimating and modeling the cure fraction in population-based cancer survival analysis. *Biostatistics*. 2007;8(3):576–94.
30. Jakobsen LH, Bøgsted M, Clements M. Generalized parametric cure models for relative survival. *Biom J*. 2020;62(4):989–1011.
31. Andersson TML, 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(1):96.
32. Jakobsen LH, et al. On estimating the time to statistical cure. *BMC Med Res Methodol*. 2020;20(1):71.
33. Romain G, et al. Time-to-cure and cure proportion in solid cancers in France. A population based study. *Cancer Epidemiol*. 2019;60:93–101.
34. Lambert PC, Dickman PW, Rutherford MJ. Comparison of different approaches to estimating age standardized net survival. *BMC Med Res Methodol*. 2015;15(1):64.
35. Reulen RC, et al. Long-term cause-specific mortality among survivors of childhood cancer. *JAMA*. 2010;304(2):172–9.
36. Brewster DH, et al. Subsequent mortality experience in five-year survivors of childhood, adolescent and young adult cancer in Scotland: a population based, retrospective cohort study. *Eur J Cancer*. 2013;49(15):3274–83.
37. Moke DJ, et al. A population-based analysis of 30-year mortality among five-year survivors of adolescent and young adult cancer: the roles of primary cancer, subsequent malignancy, and other health conditions. *Cancers (Basel)*. 2021;13(16):3956.
38. National Institute for Health and Care Excellence, Technology appraisal guidance: Olaparib with bevacizumab for maintenance treatment of advanced high-grade epithelial ovarian, fallopian tube or primary peritoneal cancer 2024.
39. National Institute for Health and Care Excellence, Single Technology Appraisal: Olaparib in combination with bevacizumab for maintenance treatment of advanced ovarian, fallopian tube and peritoneal cancer after response to first-line platinum-based chemotherapy with bevacizumab [Review of TA693] [ID4066]. Committee Papers. 2023.
40. National Institute for Health and Care Excellence, Technology appraisal guidance: Nivolumab with platinum and fluoropyrimidine-based chemotherapy for untreated HER2-negative advanced gastric, gastroesophageal junction or oesophageal adenocarcinoma 2023.
41. National Institute for Health and Care Excellence, Single Technology Appraisal: Nivolumab in combination with chemotherapy for untreated advanced gastric or gastro-oesophageal junction cancer [ID1465]. Committee Papers. 2021.
42. Guyot P, et al. Enhanced secondary analysis of survival data: reconstructing the data from published Kaplan-Meier survival curves. *BMC Med Res Methodol*. 2012;12(1):9.
43. Dickman PW, Coviello E. Estimating and modeling relative survival. *Stata J*. 2015;15(1):186–215.
44. Coviello E. *STEXPECT*: Stata module to compute expected survival, in *Statistical Software Components*. 2004, Boston College Department of Economics.
45. Lambert PC. *STANDSURV*: Stata module to compute standardized (marginal) survival and related functions. In: *Statistical Software Components*. 2021, Boston College Department of Economics.
46. Lambert PC, Royston P. Further development of flexible parametric models for survival analysis. *Stata J*. 2009;9(2):265–90.
47. Dallow N, Best N, Montague TH. Better decision making in drug development through adoption of formal prior elicitation. *Pharmaceut Stat*. 2018;17(4):301–16.
48. Gosling JP. *SHELF*: the Sheffield elicitation framework. In: Dias LC, Morton A, Quigley J, editors. *Elicitation: the science and art of structuring judgement*. Cham: Springer International Publishing; 2018. p. 61–93.
49. Lambert PC, Syriopoulou E, Rutherford MR. Direct modelling of age standardized marginal relative survival through incorporation of time-dependent weights. *BMC Med Res Methodol*. 2021;21(1):84.
50. Palmer S, et al. A guide to selecting flexible survival models to inform economic evaluations of cancer immunotherapies. *Value Health*. 2023;26(2):185–92.
51. Mounier M, et al. Flexible modeling of net survival and cure by AML subtype and age: a French population-based study from *FRANCIM*. *J Clin Med*. 2021;10(8):1657.
52. Eriksson H, et al. The proportion cured of patients with resected stage II-III cutaneous melanoma in Sweden. *Cancers (Basel)*. 2021;13(10):2456.
53. Botta L, et al. A new cure model that corrects for increased risk of non-cancer death: analysis of reliability and robustness, and application to real-life data. *BMC Med Res Methodol*. 2023;23(1):70.
54. Boulesteix AL, et al. Editorial for the special collection “Towards neutral comparison studies in methodological research.” *Biom J*. 2024;66(2): e2400031.