RESEARCH

survextrap: a package for flexible and transparent survival extrapolation

Christopher H. Jackson

Correspondence:
chris.jackson@mrc-bsu.cam.ac.uk
MRC Biostatistics Unit, University of Cambridge, Cambridge, UK
Full list of author information is available at the end of the article

Abstract

Background: Health policy decisions are often informed by estimates of long-term survival based primarily on short-term data. A range of methods are available to include longer-term information, but there has previously been no comprehensive and accessible tool for implementing these.

Results: This paper introduces a novel model and software package for parametric survival modelling of individual-level, right-censored data, optionally combined with summary survival data on one or more time periods. It could be used to estimate long-term survival based on short-term data from a clinical trial, combined with longer-term disease registry or population data, or elicited judgements. All data sources are represented jointly in a Bayesian model. The hazard is modelled as an M-spline function, which can represent potential changes in the hazard trajectory at any time. Through Bayesian estimation, the model automatically adapts to fit the available data, and acknowledges uncertainty where the data are weak. Therefore long-term estimates are only confident if there are strong long-term data, and inferences do not rely on extrapolating parametric functions learned from short-term data. The effects of treatment or other explanatory variables can be estimated through proportional hazards or with a flexible non-proportional hazards model. Some commonly-used mechanisms for survival can also be assumed: cure models, additive hazards models with known background mortality, and models where the effect of a treatment wanes over time. All of these features are provided for the first time in an R package, `survextrap`, in which models can be fitted using standard R survival modelling syntax. This paper explains the model, and demonstrates the use of the package to fit a range of models to common forms of survival data used in health technology assessments.

Conclusions: This paper has provided a tool that makes comprehensive and principled methods for survival extrapolation easily usable.

Keywords: survival; extrapolation; Bayesian; external data

1 Background

Health policy decisions are often informed by censored survival data with limited follow-up, such as clinical trial data. However, since decisions may have long-term consequences, the policy-maker is typically interested in the expected survival over

the long term. This can be difficult to estimate when the main source of data is short-term. This task is generally referred to as "extrapolation" [e.g. 28]. While this word may imply a naive assumption that short-term trends will continue in the long term, it is now widely acknowledged that making reliable decisions requires combining evidence and judgements about both the short term and the long term.

Many approaches have been suggested for combining short-term and long-term data for survival extrapolation. For reviews of these methods, see Bullement et al. [8], Jackson et al. [20], and for a broader review of flexible models for survival analysis in health technology assessments, see Rutherford et al. [34].

An overview of these approaches is now given, structured around four desirable characteristics: (a) to allow all relevant information to be included, (b) to fit the data well, (c) to allow the resulting uncertainty to be quantified, (d) to be easy to implement.

Including all relevant information As well as the short-term trial data, there may be registry data on people with the disease of interest, or national statistics describing mortality of the general population. Long-term survival can then be estimated by building a model to jointly describe all data sources [3, 13, 17], under assumptions on how the data are related, for example, proportional hazards between populations. Another common approach is based on partitioning time into different intervals in which the hazard is informed by different data [e.g. 10]. Elicited expert judgements, ideally about interpretable quantities such as survival probabilities, can also be included in survival extrapolation, either to directly provide long-term survival estimates [12] or through formal Bayesian combinations with data [11].

As well as including all relevant *data*, models should also be designed to represent any *substantive knowledge* about how risks vary with time and between people. One common assumption involves distinguishing different causes of death in the model (e.g. the cause of interest and background causes) through additive hazards [30, 31, 34]. Another commonly-modelled mechanism is the notion of "cure" [e.g 4, 9, 14], where some survivors are eventually assumed to have zero or negligible risk of some type of event in the long term. A particularly important quantity in healthcare decision models is the relative effect of a new treatment on survival, which is generally unknowable beyond the horizon of its trial, and modellers are

often advised to consider different mechanisms for how this effect might change [29].

Faithfully representing observed data A relatively easy part of this task is to estimate short-term risks from the short-term data. There is a vast range of statistical methods available for building, selecting or averaging over parametric survival models [25, 34]. This is important to do well, since the expected survival in the long term is a function of both short-term and long-term hazards. However, short-term fit is a weak guide to long-term plausibility. Extrapolations of fitted models outside the data are heavily influenced by the choice of parametric form, therefore are unreliable if this is not informed by a plausible mechanism. The most common survival models (e.g. Weibull and log-logistic) are generally chosen for their mathematical convenience and availability in software, rather than their biological plausibility [1].

Flexible parametric survival models, e.g. spline models [33] are designed to adapt their shape to fit data arbitrarily well. These allow the shape of the hazard function to change at any time, hence can adapt to fit combinations of short-term data and long-term data [17, 44]. Since the shapes of fitted spline models are driven by data, rather than knowledge about the mechanism, caution is advised when using them for extrapolation [25, 34].

Expressing uncertainty about survival An appreciation of uncertainty is important in healthcare decision-making [6]. Representing parameter uncertainty probabilistically (the Bayesian perspective) has various advantages [35] — one advantage is the ease with which multiple sources of evidence can be modelled jointly to enhance information about quantities of interest. This approach, sometimes called "multiparameter evidence synthesis", has been used for survival extrapolation [3, 9, 13, 17, 25, 31, 44]. Another advantage of the probabilistic perspective is that it allows the expected value of further information to be calculated, e.g. for longer-term follow up of survival [39, 43].

Ease of implementation For a statistical method to be useful, it should be as easy as possible to use in software. The ideal tool would allow the decision-maker to input all available data and relevant knowledge, and convert those to interesting results. Relevant assumptions should be made transparent, while the computer bears

the burden of translating knowledge and assumptions to a mathematical form and processing the data. Flexible survival models can easily be fitted to individual-level censored data, for example using the R packages `flexsurv` [21] or `survHE` [2], or the Stata package `stpm2` [27], but these do not have facilities for including "external" data to aid extrapolation. Bayesian evidence synthesis models have been implemented using Bayesian modelling languages, e.g. BUGS [17] and JAGS [44], though these require specialised statistical and programming skills.

1.1 The `survextrap` model and package

In this paper's view, there has been no method for survival extrapolation that satisfies all four of these desirable criteria. For example, while the Guyot et al. [17] model flexibly accommodates multiple sources of data, it requires specialised programming and advanced statistical expertise. The method of Che et al. [10] is based on probabilistically blending a model for short-term data with a model for long-term data, however this only accommodates two sources of data, and does not address how to develop and implement each of the two models.

This paper presents a model and R package, `survextrap`, that is intended to achieve these criteria. The model is a Bayesian evidence synthesis, that can combine right-censored individual data with any number of external data sources. The model builds on that of Guyot et al. [17] in various ways, in particular, adding the ability to supply substantive prior information on the parameters. External data are supplied in a general aggregate count form that can encompass typical population or registry data, as well as elicited judgements about survival probabilities. A penalised spline model is used that can represent hazard changes at any times. Through a Bayesian procedure, an appropriate level of smoothness and flexibility is estimated. The result is a posterior distribution that represents uncertainty about survival given all data and knowledge provided. Uncertainty about potential hazard changes in times not covered by the data can also be included in this posterior, by allowing the spline function to vary smoothly in these times. The package also implements some commonly-used mechanisms for survival extrapolation: additive hazards (relative survival) models, mixture cure models, and models where the effect of a treatment wanes over time. A model can be fitted in `survextrap` using a single call to an R function, which follows the standard syntax for survival mod-

elling in R, and a range of common summary outputs can be extracted with single function calls.

The model is fully described in Section 2, explaining the idea of M-splines, how they are used to model data, and their extensions to deal with explanatory variables and special mechanisms. Section 3 introduces the `survextrap` package and points to an example of its basic use. Section 4 demonstrates how the model and package might be used in a realistic application, to an evaluation of the survival benefits of cetuximab in head and neck cancer [17]. A range of models are compared, each with different data sources and assumptions about how inferences outside the data are made. The discussion (Section 5) gives some suggestions for further research.

2 Methods: statistical model

We suppose there is:

(a) an individual-level dataset, with survival times that may be right-censored,

(b) optionally, also one or more aggregate external datasets, giving counts of survivors over arbitrary time periods.

The external datasets, indexed by $j = 1, \ldots, J$, take the following form:

- out of n_j people alive at a time u_j, with common characteristics $\mathbf{x}_j$,
- r_j of them survive until the time v_j.

This form of data might be derived from e.g. disease registries or population life tables. It may also be derived from expert elicitation of the survival probability p_j over the period (u_j, v_j), using the following method. Suppose we elicit a $Beta(a, b)$ prior for p_j. We interpret this as the posterior distribution from a vague prior (Beta(0,0), say) for p_j combined with data (r_j, n_j), which would be a $Beta(r_j, n_j - r_j)$. Then, by equating $a = r_j, b = n_j - r_j$, we can deduce the data (r_j, n_j) that would represent knowledge equivalent to the elicited judgement. See Cooney and White [11] for a related approach.

A single model is assumed to jointly generate all sources of data. This is defined by its hazard function $h(t|\boldsymbol{\theta}, \mathbf{x})$, where t is time, $\boldsymbol{\theta}$ includes all parameters, and $\mathbf{x}$ includes predictors (e.g. characteristics of individuals, or variables that distinguish one dataset from another). This model will be based around a flexible function known as an M-spline [32], as previously used by Brilleman et al. [7] and Król et al. [26] for survival modelling. M-splines have some computational advantages over

other kinds of splines and flexible models, as discussed in Appendix (A). Appendix (B) explicitly describes the differences between the M-spline model and the cubic spline model used by Guyot et al. [17].

Sections 2.1–2.3 describe the details of the M-spline model. This core model can then form the basis of some other specialised survival modelling mechanisms: additive hazards, cure models and treatment waning, as described in Section 2.4. The model will be fitted by Bayesian inference (Section 2.5), which produces a posterior distribution for the parameters $\boldsymbol{\theta}$. Estimates and measures of uncertainty for long-term survival and other quantities of interest can then be deduced.

2.1 M-spline model

In an M-spline model, the hazard $h(t)$ at time t is defined by a weighted sum of *basis functions*, which takes the form:

$$h(t) = \eta \sum_{i=1}^{n} p_i b_i(t)$$

The *scale* parameter η is proportional to the typical level of the hazard, and the *basis coefficients* $p_1, \ldots, p_n$ satisfy $\sum_i p_i = 1$. A simple example is illustrated in Figure 1 with $n = 5$ basis functions. The axis of time is split into regions defined by $n - 1$ "knots", in this example at 5, 10, 15 and 20. Given these knots, a set of n basis functions $b_i(t)$ are defined to span the range of the knots, where the first one has a peak at zero, the next $n - 2$ functions have peaks inside the knots, and the final basis function is constant when t exceeds the final knot. The basis functions are polynomials (cubics by default), restricted to be positive and to ensure that $h(t)$ is a smooth function of t. The coefficients that represent a constant hazard function can be derived as a deterministic function of the knots. The full definition is given in Appendix (C).

The parameters η and p_i do not have exact interpretations — the intention is to obtain a function that adapts to fit all available data, rather than to learn the biological or clinical mechanism. Three examples of M-spline hazard functions, on the same basis and scale, but with different coefficients, are illustrated in Figure 1.

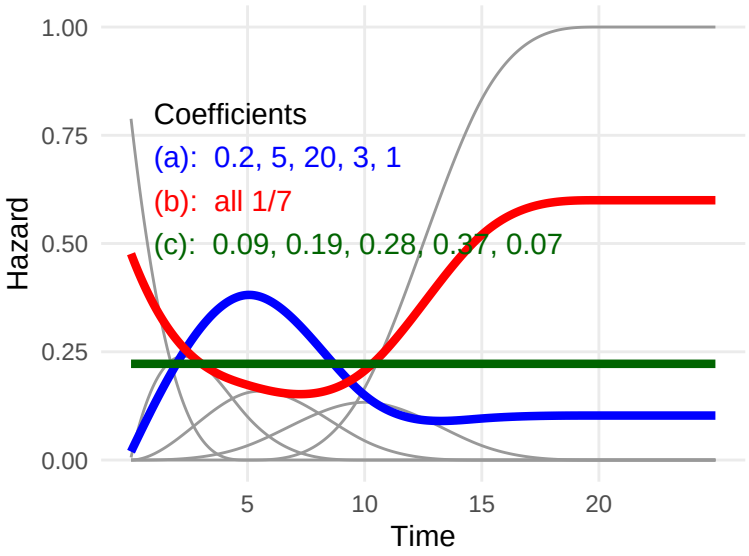

Figure 1 Three theoretical examples of M-spline hazard functions $h(t) = \sum_i p_i b_i(t)$ (coloured lines). Each is defined as a linear combination of the same 5 polynomial basis functions $b_i(t)$ (in grey), and scale $\eta = 3$. Each has different basis coefficients p_i, indicated in the plot. Note that the constant hazard model is not the one where all the coefficients are the same.

2.2 Modelling data with an M-spline model

The knots should be chosen to allow the hazard function to be determined from the data, and to allow the shape of the function to change outside the data, if this is plausible and an estimate outside the data is needed. This section explains the default approach in the `survextrap` package, but these defaults can be modified by placing knots anywhere if needed.

As in Royston and Parmar [33], knots are placed by default at quantiles of the uncensored individual-level survival times. If there are also external data, additional knots should be defined by the user if required to cover the times of these data. The appropriate number and location of additional knots depends on how many external datasets there are and what times they cover — noting that hazard changes *within* an interval (u_j, v_j) cannot be identified from an aggregate count of survivors over this interval. Then to allow for hazard changes outside the period covered by all data, the highest knot should be placed at a point beyond which either we assume the hazard will not change, or any hazard changes are unimportant.

The appropriate level of flexibility for the hazard function is determined automatically from the data, by using a principle of *penalisation* [46], or "shrinkage"

towards a constant hazard. Firstly, the number of knots spanning the individual data is fixed to be large enough to accommodate all plausible shapes ($n = 10$ basis functions is the current package default). A hierarchical prior is then placed on the coefficients p_i, to represent a belief that the true model *may* be this flexible, but is most likely to be less flexible. Then, when the model is fitted to data, the resulting posterior represents the optimal level of flexibility needed to describe the data. This is intended to protect against the risk of over-fitting.

Specifically, the prior for the p_i is a multinomial logistic distribution: $\log(p_i/p_1) = \gamma_i$, with $\gamma_1 = 0$ and $\gamma_i \sim Logistic(\mu_i, \sigma)$ for $i = 2, \ldots, n$. The prior mean $\boldsymbol{\mu} = (\mu_2, \ldots, \mu_n)$ is defined so that the corresponding p_i represent a constant hazard $h(t)$. The prior variability σ controls the smoothness of the fitted hazard curve, and is estimated from the data. If $\sigma = 0$ then we are certain that the hazard is constant, and values of σ around 1 favour "wiggly" curves where the hazard is driven mainly by the data. See the `survextrap` package vignettes for some examples.

2.3 Modelling explanatory variables

To extend this model to allow the hazard $h(t|\mathbf{x}) = \eta \sum_i p_i b_i(t)$ to depend on explanatory variables $\mathbf{x}$, a proportional hazards model can be used, where the scale parameter η is redefined as $\eta(\mathbf{x}) = \eta_0 \exp(\boldsymbol{\beta}^T \mathbf{x})$.

A novel, flexible non-proportional hazards model is also defined, by also modelling the spline coefficients p_i as functions of $\mathbf{x}$ using multinomial logistic regression:

$$\begin{aligned} \log(p_i(\mathbf{x})/p_1(\mathbf{x})) &= \gamma_i(\mathbf{x}) \\ \gamma_i(\mathbf{x}) &\sim Logistic(\mu_i + \boldsymbol{\delta}_i^T \mathbf{x}, \sigma) \quad (i = 2, ..., n), \quad \boldsymbol{\delta}_1 = 0 \end{aligned}$$

The sth element of the vector $\boldsymbol{\delta}_i$ describes the departure from proportional hazards for the sth covariate in the region of time associated with the ith spline basis term. With all $\boldsymbol{\delta}_i = 0$, this is the proportional hazards model. For each covariate s, a hierarchical prior is used for the non-proportionality effects, $\delta_{is} \sim Normal(0, \tau_s)$, which "partially pools" or smooths the effects from different regions of time i.

Hence, in the non-proportional hazards model, the ratio between the hazards $h(t|\mathbf{x})$ at different covariate values $\mathbf{x}$ is a function of time t. Since the $\boldsymbol{\delta}_i$ may be arbirtarily different in each region of time i, the hazard ratio can be an arbitrarily-

flexible function of time. The shape of this time-dependence is estimated from the data, while the hierarchical prior protects against over-fitting.

2.4 Special mechanisms

The idea behind the package is for inferences to be driven by transparently-stated data and judgements. As described so far, the flexible model used to achieve this is treated as a "black box", that is, its specific form is not intended to have a biological or clinical interpretation. However, sometimes plausible mechanisms can be specified via parametric model structures, to improve estimates of long-term survival. The package currently supports the following three mechanisms, that are sometimes assumed for survival extrapolation.

Additive hazards / relative survival Here the overall hazard is assumed to be the sum of two hazards for different causes of death, as $h(t) = h_b(t) + h_c(t)$, where

- $h_b(t)$ is the "background" hazard, assumed to be a known, piecewise-constant function of time. This is often confidently known from national statistics on general mortality.
- $h_c(t)$ is the "cause-specific" or "excess" hazard for the disease of interest, which is modelled with the M-spline parametric model.

The individual-level data are assumed to be described by $h(t)$, and fitting the model to these data involves estimating the parameters governing $h_c(t)$. The corresponding survivor functions are multiplicative: $S(t) = S_b(t)S_c(t)$, hence the alternative term "relative survival" model. This is a variant of the model used by Nelson et al. [30], but Bayesian and with a different kind of spline.

Mixture cure model In a mixture cure model, data are assumed to arise from a survival function $S(t|p, \boldsymbol{\theta}) = p + (1 - p)S_0(t|\boldsymbol{\theta})$, where the unknown parameter p is sometimes termed the "cure probability", and $S_0(t|\boldsymbol{\theta})$ is a parametric model with parameters $\boldsymbol{\theta}$, termed the "uncured survival". Here, S_0 is the M-spline model described above. For very large t, the survival converges to a constant p, the probability of never experiencing the event. This is often interpreted as a mixture of two populations, where a person has zero hazard at all times t with probability p, or a hazard $h_0(t)$ at all times with probability $1 - p$. However, contrary to some descriptions of this model, this is not a necessary assumption, because the same

survival function arises if everyone is subject to the same hazard that decreases asymptotically to zero over time, $h(t) = f_0(t)/\left(p/(1-p) + S_0(t)\right)$, where $f_0(t)$ is the probability density function of the "uncured" model. These two interpretations are indistinguishable in practice, since in the mixture interpretation, we cannot determine which of the two populations censored observations belong to.

A mixture cure model can either be used for the overall hazard, or the cause-specific hazard h_c in an additive hazards model. Using a cure model for h_c would be appropriate if we can assume that the cause-specific hazard decreases to a negligible amount over time, so that everyone with the disease of interest is essentially "cured", and their hazard becomes dominated by background causes.

Waning treatment effects Health technology assessments are often primarily informed by trials of a novel treatment against a control. Beyond the trial horizon, information about the relative effect of the new treatment will be weak. A naive extrapolation from the model would assume that the estimated short-term effect will continue in the long term (e.g. as a constant hazard ratio). This is often contrasted with more conservative scenarios where the treatment effect wanes over time, so that after some point, treated and control patients have the same hazard.

Treatment effect waning can be achieved by firstly fitting a parametric model $h(t|x)$, including the effect of a treatment x, to short-term data. This does not necessarily need to be a proportional hazards model. Then, the predicted hazard for the control group is taken from the fitted model, $h(t|x = 0)$. The predicted hazard for the treated group is obtained as $h(t|x = 1) = h(t|x = 0)hr(t)$, where the time-dependent hazard ratio $hr(t)$ is defined as follows:

- For $t \leq t_{min}$, $hr(t)$ is taken from the fitted model. t_{min} might be the end of the trial follow-up period, or a later point up to which the effect from the trial can be confidently extrapolated.
- For $t \geq t_{max}$, $hr(t) = 1$.
- For $t_{min} \leq t \leq t_{max}$, $\log(hr(t))$ is assumed to diminish linearly between $\log(hr(t_{min}))$ at t_{min}, and zero at t_{max}.

2.5 Bayesian inference

The models define a hazard function $h(t|\boldsymbol{\theta}, \mathbf{x})$, from which the cumulative hazard function and survivor function $S(t|\boldsymbol{\theta}, \mathbf{x})$ can be derived, to construct the likelihood

function for the individual-level data. This hazard function is also assumed to govern the external datasets j, with any differences between them explained by different explanatory variables $\mathbf{x} = \mathbf{x}_j$ and the time period covered. The likelihood for each external dataset j is built by assuming that r_j is generated from a Binomial distribution with denominator n_j and probability $p_j = S(u_j|\boldsymbol{\theta}, \mathbf{x}_j)/S(t_j|\boldsymbol{\theta}, \mathbf{x}_j)$, that is, the probability of survival to time u_j conditionally on being alive at t_j. Samples from the posterior distribution of $\boldsymbol{\theta}$ (which may comprise e.g. η, $p_1, \ldots, p_n$, σ, $\boldsymbol{\beta}$, $\boldsymbol{\delta}$, τ_s, depending on the model choice) are then obtained by Markov Chain Monte Carlo methods, specifically Hamiltonian Monte Carlo as implemented in Stan [37].

In the package, all priors for the parameters comprising $\boldsymbol{\theta}$ can be defined by the user. If any priors are not specified, then the following defaults are currently used. These are "weakly informative", that is, containing some stabilising information but largely letting the data drive the inferences, following principles described by Gelman et al. [15]. As in Brilleman et al. [7], the baseline log hazard $\log(\eta_0)$ (for covariate values of zero) is given a normal prior with mean zero and standard deviation 20. For the log hazard ratios, Normal(0, 2.5) is used, and a Gamma(2,1) is used for the smoothness parameter σ.

Prior calibration Procedures are also provided for simulating from the joint prior distributions for the parameters in $\boldsymbol{\theta}$, to ensure that they imply plausible beliefs about easily-understandable quantities. For example, σ governs how much the hazard is expected to vary through time — a constant hazard has $\sigma = 0$, but other values of σ are hard to interpret. However, we can draw a simulated value from any given prior for σ, jointly with the the distributions for the p_i and η, and deduce the implied hazard curve $h(t)$. The hazard variability in this curve could be described by the ratio ρ between (say) the 90% percentile and 10% percentile of $h(t)$ over a fine, equally-spaced grid of times t from zero to the highest knot. By repeatedly simulating hazard curves, we can draw from the prior distribution of ρ. We can then calibrate the prior for σ to achieve a prior on ρ that expresses beliefs of the form "the highest values of the hazard are unlikely to be 10 times the lowest values".

Statistical model comparison The goodness-of-fit of different models to the observed data can be compared using leave-one-out cross-validation, via the method and R package of Vehtari et al. [40, 41]. For each observation i (individual event or

censoring time in the individual data, or individual event indicator in the external data), this method estimates $elpd_i$, the expected log predictive density, a measure of the accuracy with which a model would predict the ith observation if it were left out when fitting the model. The sum LOOIC $= -2\sum_i elpd_i$ is then used as an "information criterion" to compare the fit of models. LOOIC is similar in principle to DIC [36], but with a direct interpretation in terms of predictive ability.

3 Implementation of the software

An R package called `survextrap` implements the method. It is available from `https://chjackson.github.io/survextrap/`. It uses the `rstan` interface to the Stan software [37, 38] to perform Hamiltonian Monte Carlo sampling from the posterior distribution of the Bayesian model. Models can be fitted with a single R command, using a similar mechanism to the `rstanarm` package [7]. Posterior summaries (e.g. estimates and credible intervals) for a range of interesting outputs (e.g. survival probabilities, hazards, mean survival times) can then be extracted using single commands. Outputs from all these functions obey "tidy data" principles [45], to facilitate further processing, in particular, plotting with the `ggplot2` R package.

An example of basic use of the package is given in Appendix (D) of this paper. The website `https://chjackson.github.io/survextrap/` gives thorough documentation for all the package's functions, including a series of articles describing specific features in more detail, and the code and data to reproduce the analysis in Section 4.

4 Demonstration: cetuximab for head and neck cancer

This section demonstrates a range of models that can be built with `survextrap` in a typical application to a health technology evaluation, each using different kinds of information and model assumptions to enable extrapolation. The full R code to reproduce each model fit, calculation and plot is supplied and explained in a supplementary article. The paper will focus on discussing different analysis choices and their consequences.

We study the data that were previously analysed by Guyot et al. [17], describing the survival of people with head and neck cancer. Data are obtained from 5 years of follow-up of a trial [5] of cetuximab and radiotherapy, compared to a control group who only received radiotherapy. Individual-level data were imputed from the

published Kaplan-Meier estimates, using the method of Guyot et al. [16]. As well as the trial data, there are two external data sources [full details are given in 17]:

(a) a cancer registry (SEER: the Surveillance, Epidemiology and End Results Program), representing people with the same distribution of age, sex, cancer site and calendar period of diagnosis date as the trial data. This gave counts of survivors r_j over the following year, out of n_j alive at j years after diagnosis, for $j = 5$ to 25, giving estimates of annual survival probabilities p_j.

(b) survival data for the general population of the United States, matched by age, sex and date.

We examine either the survival for the control group alone, or the increase in survival provided by cetuximab. Specifically we calculate the restricted mean survival time (RMST), or the difference in restricted mean survival times, over time horizons varying from 5 years up to 40 years. We discuss how longer-term data and judgements are required to obtain confident estimates of longer-term survival.

4.1 Prior information

To improve transparency, and stabilise computation, we specify priors explicitly, rather than relying on the package's (fairly vague) defaults.

1 Hazard scale parameter η. Since patients in the trial have a median age of 57 (range 34 to 83), the prior for η is calibrated to imply a prior mean survival of 25 years after diagnosis but with a variance chosen to give a wide 95% credible interval of about 5 to 100 years. Note that this interval describes *uncertainty* about knowledge of the *mean* in the control group, not *variability between individuals* in this group.

2 Smoothing parameter σ. This is chosen by simulation (as described in Section 2.5) so that the highest hazard values for the control group (90% quantile) are expected to be $\rho = 2$ times the lowest values (10% quantile), but with a credible interval for ρ of between 1 and 16.

The individual-level data are spanned by $n = 6$ spline basis terms, chosen according to the quantiles of observed event times. Beyond the trial data, additional knots are used depending on what external data are included, and what time horizon we want to estimate survival for.

4.2 Trial data alone: extrapolating a single arm

Firstly we study just the data from the control treatment group. We contrast two models that describe the trial data in the same way, but differ in how extrapolation is done, labelled as:

(1a) the highest knot is set to 20 years,

(1b) the highest knot is set to the final event time in the data (5 years in this case). This is the package default if an upper knot is not specified.

The posterior distributions of the survival and hazard curves up to 20 years (Figure 2) show how extrapolation relies on both data and assumptions. Here there are no data describing 5 to 20 years. In (1b) we made the strong assumption that hazards will remain constant after the trial horizon of 5 years. In (1a) we assumed that the hazard will change smoothly after the trial, but using a spline model that allows any size and direction of change, not determined by the fit to the short-term data. Therefore there is a lot of uncertainty around the extrapolated hazard function in (1a), but the extrapolation under (1b) is more confident. The exact extent of uncertainty in (1a) will be sensitive to where knots are placed, though a rough uncertainty quantification may be sufficient to highlight the need for further information beyond that included in the trial.

5 and 20-year RMST estimates are shown in Table 1. Credible intervals for 20-year estimates are wide when we do not constrain the extrapolated hazard. The 5 year RMSTs from model (1a) did not change by more than 0.1 years when the model was made more or less flexible (through the number of basis terms varying from 5 to 12).

4.3 Trial data alone: treatment comparisons

Now we consider a comparison between treatment groups based on the trial data alone, using three alternative models, labelled as:

(2a) a proportional hazards model,

(2b) a parametric non-proportional hazards model (Section 2.3),

(2c) both treatment arms modelled separately.

These models fit similarly well to the trial data, judging from the fitted survival curves (Figure 3), and their similar LOOIC cross-validation statistics (Table 2). The difference between them is more apparent when extrapolating. The upper boundary

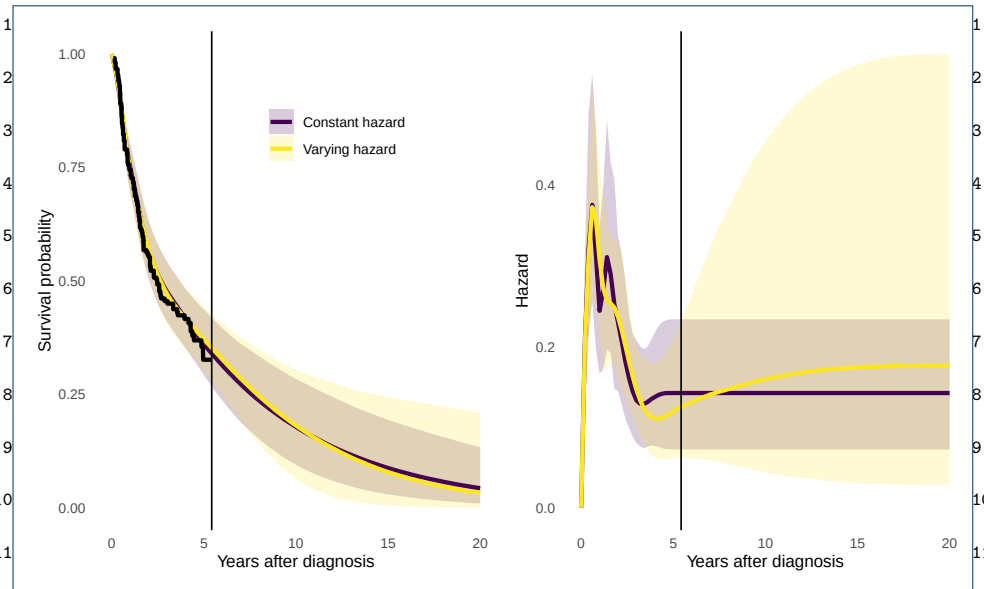

Figure 2 Survival and hazard estimates (posterior median and 95% credible intervals) for the trial control group, from flexible Bayesian models where the hazard after the last event time in the trial (vertical line) is either arbitrarily varying (1a) or constant (1b). Kaplan-Meier estimate from the trial data also shown on the left as a black jagged line.

knot is set to 20 years, so that we allow the hazard to change after 5 years, even though there is no data then. Over five years (Table 2) the survival and incremental survival between treatment groups is similar between the three models, but over 20 years the uncertainty about these quantities is greater. The credible intervals are narrowest under the proportional hazards model, and widest when modelling arms independently. The non-proportional hazards model makes more efficient use of the data than modelling arms separately, though the proportional hazards model is adequate (judging by LOOIC).

All the models so far have ignored the substantive information that exists beyond the trial data: the registry and population data to inform mortality for these patients, and information about the mechanism of the treatment effect.

4.4 External data from the patients of interest

We now examine how to incorporate external data in `survextrap` models. Annual hazard (mortality rate) estimates from the SEER registry data, calculated as $-\log(r_j/n_j)$, are illustrated in Figure 4, with corresponding interval estimates (calculated from quantiles of the $Beta(r_j, n_j - r_j)$). These are included in a joint model with the trial data (labelled (1c) in Table 1), with knots added at 10, 15

Model	Observed data	Extrapolation assumptions	Time horizon	Restricted mean survival
(1a)	Trial	No extrapolation	5	2.88 (2.63, 3.13)
(1a)	Trial	Uncertain hazard	20	5.11 (3.77, 7.14)
(1b)	Trial	Constant hazard	20	5.12 (4.01, 6.71)
(1c)	Trial,registry	No extrapolation	20	5.76 (5.04, 6.57)
(1d)	Trial,registry	Uncertain hazard	40	6.2 (5.35, 7.12)
(1e)	Trial,registry,population	Uncertain excess hazard	40	6.22 (5.36, 7.13)
(1f)	Trial,registry,population	Mixture cure	40	6.27 (5.39, 7.22)
(1g)	Trial,registry,population	Elicited survival	40	6.26 (5.37, 7.17)

Table 1 Comparison of estimates of restricted mean survival time in years (posterior median and 95% credible intervals) over different models and time horizons, for head and neck cancer patients in the control group. The models differ by the different sources of observed data included, and the different assumptions used for extrapolation outside the time horizon of the observed data.

Model	Restricted mean survival (control)	Increase in mean survival (cetuximab - control)	LOOIC
Trial data alone: prediction horizon 5 years			
(2a) Proportional hazards	2.89 (2.62,3.14)	0.31 (-0.06,0.67)	1156
(2b) Non-proportional hazards	2.88 (2.62,3.14)	0.31 (-0.06,0.68)	1157
(2c) Separate arms	2.88 (2.63,3.13)	0.36 (0,0.73)	1160
Trial data alone: prediction horizon 20 years			
(2a) Proportional hazards	4.97 (3.82,7.11)	1.1 (-0.22,2.61)	1156
(2b) Non-proportional hazards	4.98 (3.76,6.66)	1.12 (-0.91,3.26)	1157
(2c) Separate arms	5.11 (3.77,7.14)	1.33 (-1.24,4.12)	1160
Trial and registry data: prediction horizon 20 years, proportional hazards models			
(2d) No waning	5.82 (5.07, 6.62)	1.23 (-0.34, 2.86)	
Trial, registry and population data: prediction horizon 20 years, proportional hazards models			
(2e) No waning	5.89 (5.11, 6.65)	1.08 (-0.4, 2.61)	
(2e) 5 to 20 years	5.89 (5.11, 6.65)	1.02 (-0.38, 2.47)	
(2e) 5 to 6 years	5.89 (5.11, 6.65)	0.81 (-0.3, 1.92)	

Table 2 Comparison of models fitted to both treatment and control trial data. Estimates of restricted mean and increase in restricted mean survival in years over 5 or 20 year horizons (posterior median and 95% credible intervals), and LOOIC model comparison statistic (lower indicates better predictive ability)

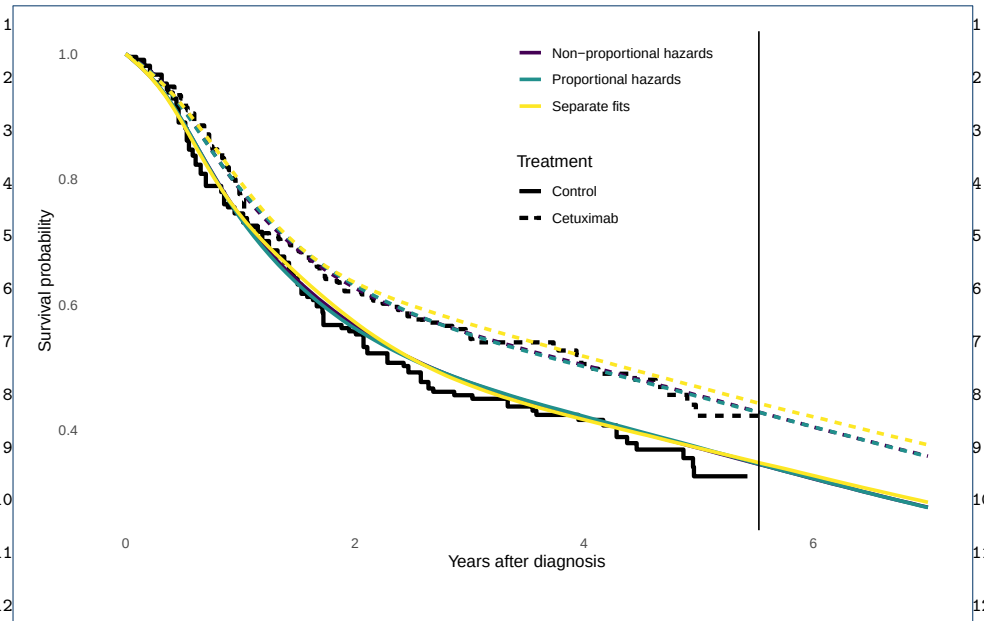

Figure 3 Estimates of survival as posterior medians from three different flexible Bayesian models for both treatment groups. Kaplan-Meier estimates from the trial data are shown as jagged lines, and spline knot locations as vertical lines.

and 20 years, and the patients in the registry are assumed to have the same survival as the control group of the trial. The posterior distribution of the hazard from this model is also illustrated in Figure 4, along with estimates from the equivalent model (from Figure 2) that excludes the registry data. The registry data makes the extrapolated hazard and RMST much more confident. The model allows the hazard to vary flexibly up to 20 years, and those variations can be estimated from the registry data. Different knot placements did not substantially affect estimates of survival over 20 years or improve the fit to the external data as measured by LOOIC (see the supplementary material for more details).

Modelling differences between external and trial data Note that in the `survextrap` package, the external data does not have to describe a population identical to that described by a particular subgroup of the individual data. The differences between data sources could instead be explained by covariates included in the model, using either proportional or non-proportional hazards [13].

The simplest model of this kind would have a covariate that indicates the dataset: e.g. a binary variable taking values "trial" and "external". To estimate the hazard ratio between datasets, we would then need data from the same time period to be

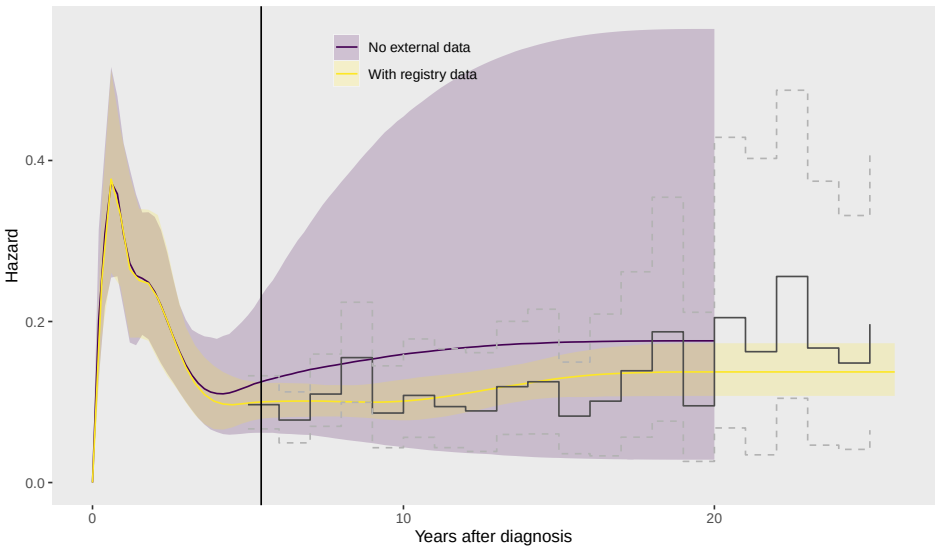

Figure 4 Hazard estimates (posterior median and 95% credible intervals) from models for trial data with (1c) and without (1a) external data from the cancer registry (coloured bands), and hazard estimates from the registry data alone (jagged lines)

observed in both datasets, or strong prior information. This is not available in the data provided with Guyot et al. [17], where the external data starts where the trial data ends at 5 years (though Figure 4 gives weak evidence that the hazard at 5 years is roughly the same in both datasets). Furthermore, to use a fitted model of this kind to predict outside the data, we would also need to assume that the relation between the datasets (e.g. proportional hazards) holds outside the observed period, which would not be verifiable from the data. We would also need to specify whether to predict for the trial or external population.

If further covariates are recorded in both datasets, these may also be included in the model to explain any differences between the datasets. Then, in theory, we may use the fitted model to make predictions for populations defined by combinations of the trial and external populations. Though likewise, extrapolation would rely on assuming that estimated covariate effects are valid outside the time and population that they were estimated from. Any combination of data from different sources needs careful consideration of potential biases.

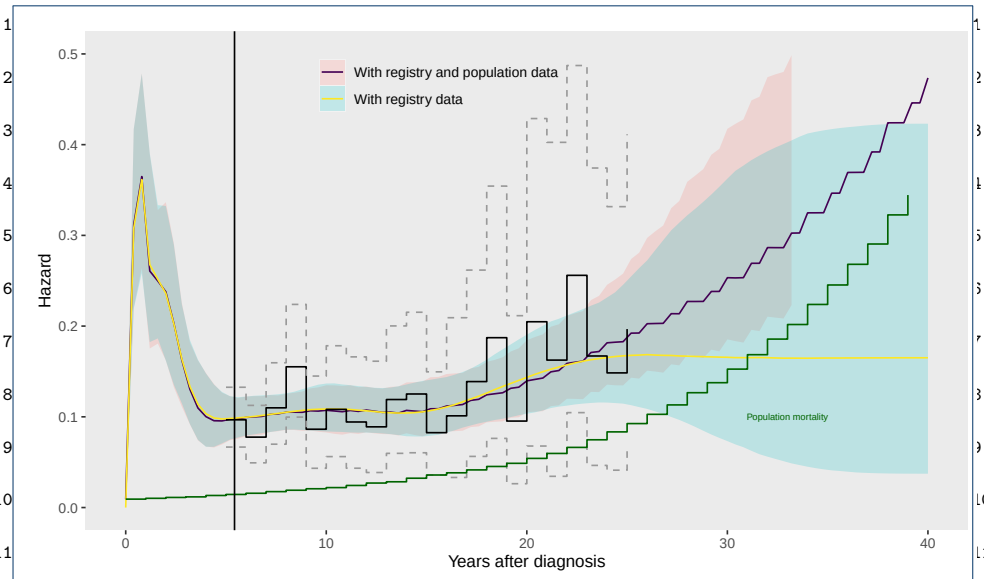

Figure 5 Hazard estimates (posterior median and 95% credible intervals) from models from the trial data and cancer registry data, compared without (1d) and with (1e) additional data from the general population.

4.5 Population data informing background mortality

Another way of including external data is through additive hazards, as described in Section 2.4. Here this allows the data on survival of the general population to be included. These are assumed to follow the background hazard $h_b(t)$, which is assumed known. The trial data follow the overall hazard $h(t)$, and the excess hazard $h_c(t)$ for head and neck cancer patients is assumed to follow the flexible M-spline model and estimated. This model constrains the survival of head and neck cancer patients to be no better than the survival of the general population.

The population data are added to the model that includes the registry data. We compare hazard extrapolations up to 40 years, placing further knots at 30 and 40 years (in addition to those spanning the trial and registry data), either without or with the population data (Figure 4), labelled (1d) and (1e) in Table 1. The population data do not affect the hazard estimates up to 20 years, but the extrapolations over 40 years are very uncertain unless the population data are included. Including the population data allows the reasonable constraint that hazards will not go below those of the general population. The exact excess risk for head and neck cancer patients is still uncertain, however, since we do not have data beyond 25 years to inform it.

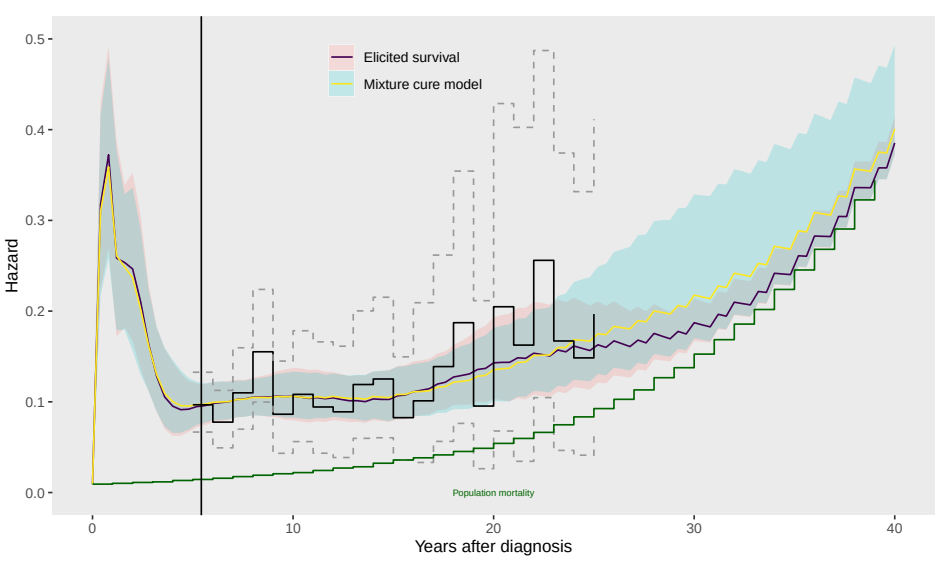

Figure 6 Hazard estimates (posterior median and 95% credible intervals) from models from the trial, cancer registry and general population data under two alternative models ((1f) mixture cure, and (1g) elicited survival) that represent a decreasing excess hazard for people diagnosed with head and neck cancer.

4.6 Mixture cure model

We could improve the precision of the estimates of the excess hazard for head and neck cancer patients by including judgements, for example, that the excess hazard will diminish to zero as people age. While there is no evidence of this from the registry data in this example, in some applications it might be plausible. One way to represent this might be through a mixture cure model (Section 2.4) fitted to the trial, registry and population data combined. Comparing the results from the mixture cure model (Figure 6, and (1f) in Table 1) to the model for the same data with no cure assumption (Figure 5 and (1e)) shows how the assumption of cure has pulled the hazard extrapolations for 20-40 years closer to the estimates of the background hazard, though with wide credible intervals. The exact shape of the extrapolation for the cure model is influenced by the parametric form for the mixture cure hazard function. In practice, this should be checked for plausibility.

4.7 Elicitation of long-term survival probabilities

A more flexible way to include longer-term judgements is by eliciting survival probabilities. These can be converted to artificial datasets represeting counts of survivors,

which can be included as "external data" in the model, using the idea described in Section 2.

For example, we could state an assumption of "cure" in the form: "by 40 years after diagnosis, we are confident that the patients of interest will have similar mortality to the background population". The annual survival probability in the matched general population dataset at this point is 0.72. Suppose we then elicited a $Beta(1000 \times 0.72, 1000 \times (1-0.72))$ distribution, $Beta(724, 276)$, which has a 95% credible interval of (0.70, 0.75). This is equivalent to having observed $r_j = 724$ survivors at the end of the year, from $n_j = 1000$ people alive at the start. The denominator n_j could be controlled to give different amounts of prior uncertainty, e.g. $n_j = 100$ would give a $Beta(72, 100 - 72)$ which has a wider credible interval of $(0.63, 0.80)$.

This artificial dataset is then concatenated with the SEER registry data, and supplied to the model in the same way as the registry data. The predicted hazard from this model is shown in Figure 6. This reflects our assumption that the overall hazard approaches the background hazard at 40 years, but with some uncertainty. This model makes less restrictive parametric assumptions than the mixture cure model — since spline knots are placed at 20, 30 and 40 years, the hazard curve is allowed to change within a wide variety of smooth shapes. The assumption that the cancer patients are "cured" by 40 years is provided through a directly-stated judgement about survival at 40 years, rather than through extrapolating a parametric equation estimated only from data up to 25 years.

Finally, note that the RMST estimates (Table 1) do not change much between the four different assumptions (1d)–(1g) used for estimating the hazard between 20 and 40 years, since the probability of survival beyond 20 years is low.

4.8 Waning treatment effects

We have now built in a model that includes all background information about general mortality of the patients in the trial, allowing us to extrapolate absolute survival of the control group as confidently as the data allow. In Section 4.3 we built models to estimate treatment effects from the trial data. We now consider what judgements might be made about the treatment effect beyond the trial horizon, and how these can be modelled with `survextrap`.

As discussed by Guyot et al. [17], the mechanism of cetuximab is to enhance the effect of radiotherapy. The effects of both of these therapies is expected to be limited to the initial 5 or 6 years, where most of the mortality due to the cancer occurs. Therefore Guyot et al. [17] judged that the hazard ratio for the effect of adding cetuximab to radiotherapy is expected to diminish to one by around 6 years, though acknowledged some uncertainty around this.

Therefore the model which includes registry and population data but no cure is extended to include a proportional hazards model for treatment (the treatment effect mechanism that best fitted the trial data in Section 4.3). The results from this model are labelled (2e) in Table 2. The incremental restricted mean survival over 20 years is compared between three different assumptions about the treatment effect beyond the 5-year trial horizon:

(a) the log hazard ratio remains constant at the value estimated from the trial
(b) the log hazard ratio wanes linearly from the trial value at 5 years to zero at 6 years
(c) the log hazard ratio wanes linearly from the trial value at 5 years to zero at 20 years

Using all three data sources, with no waning, the incremental restricted mean survival over 20 years is estimated as 1.08 (-0.4, 2.61). This reduces to 1.02 (-0.38, 2.47) when waning is applied gradually from 6 to 20 years, and even further to 0.81 (-0.3, 1.92) when the effect is assumed to wane rapidly from 5 to 6 years. Note also that omitting the population data from this model ((2d) in Table 2) impacts the estimated treatment effect.

The assumptions made here are uncertain, and there are many ways in which this uncertainty could be described. A simple deterministic sensitivity analysis is done here, which has an advantage of transparency to decision-makers. An alternative approach would be to represent this uncertainty probabilistically (see, e.g Guyot et al. [17] for one approach), though formally specifying and eliciting distributions for a weakly-informed, time-varying quantity like this is challenging in general.

5 Discussion

This paper has introduced a tool that makes principled methods for survival extrapolation straightforward. It accommodates a wide range of data sources, that can be

represented in a flexible statistical model. The Bayesian approach allows uncertainty to be quantified, and the R package removes the need for specialised programming. The model can represent uncertainty about how the hazard will change beyond the data, assuming only that the hazard function is smooth.

While the model is flexible, all models are based on assumptions. The package tries to make these as transparent as possible. In particular, prior distributions can easily be chosen to represent beliefs about interpretable quantities. While the spline model relies on a choice of knots, the statistical fit of different choices to data can be compared. Extrapolations outside data may be sensitive to modelling choices, but uncertainty is inevitable when data is weak. If there is uncertainty, there is a tension between decision-making and recommending collection of further data. The Bayesian approach represents uncertainty using probability distributions, which allows the use of "value of information" methods to estimate the expected benefits of further information (e.g. from a health economic perspective). In principle, the posterior distribution from a `survextrap` model might be used to calculate the expected value of sample information for a new trial or further follow-up from an existing trial — the implementation details have not been worked out, but see e.g. Vervaart et al. [42] for a potential starting point.

There are several ways in which the model used here might be extended. *Hierarchical* or random effects models are one potential direction, as in `rstanarm` [7]. These might be used to represent various kinds of heterogeneity in survival, e.g. between observed groups such as different hospitals [19], or between latent classes of individuals [14]. Survival models with random effects can also be used for (network) meta-analysis [23]. Another common extension of survival models is to *multi-state* models for times to multiple events. See, e.g. Jackson et al. [22] for a comparison of flexible parametric frameworks for multi-state models, and Jansen et al. [24] for network meta-analysis of survival data with multiple outcomes. The ideas described in this paper would enable any of these previous methods to be strengthened by including background information from external data.

A final point to consider is that getting new statistical methods into routine practice involves several "phases" of research [18]. This paper has described the theoretical basis for a novel method, shown its utility in a realistic application, and provided software to make it usable with the minimum of effort. However, flexi-

ble Bayesian evidence synthesis methods are complex and specialised. To improve confidence in them, more work to demonstrate their use in a wide range of applications would be helpful. Furthermore, constructing the flexible model relies on many assumptions made for mathematical or computational convenience, such as the M-spline structure, smoothing procedure and default priors. Simulation studies would be valuable to assess whether the default models give estimates that are reliable in realistic situations, in particular when compared to better-known models. Further work on constructing practicable, flexible models that can reflect biological or clinical mechanisms, or models with stronger theoretical optimality properties, would also be beneficial. Education in statistical skills is also important. The Bayesian spline models used here are more complex than basic parametric survival models, with many ingredients that may be unfamiliar, such as prior distributions, spline knots and Markov Chain Monte Carlo computation. The online documentation includes lots of worked examples to explain the important concepts, and will be updated as a "live" resource in response to users' needs if the package becomes more widely used.

Declarations

Ethics approval and consent to participate Not applicable. Only openly-available data are studied.

Consent for publication Not applicable. Only openly-available data are studied.

Availability of data and materials All data analysed during this manuscript are made available inside the `survextrap` package. A detailed article explaining the case study in section 4, with embedded R code to directly reproduce all results including graphs and tables, is available in the supplementary file `cetuximab.html`, and in a "live" version at `https://chjackson.github.io/survextrap/articles/cetuximab.html` which will keep it up to date with any future enhancements or fixes to the software.

Competing interests The author declares that they have no competing interests.

Funding Funding was from the Medical Research Council, programme number MRC_MC_UU_00002/11. The funding body had no role in the conduct of the work.

Authors' contributions All work for the manuscript was undertaken by Christopher Jackson.

Acknowledgements I am grateful to Iain Timmins for suggesting the smoothness constraint at the spline boundary, and other helpful discussion. Thanks also to Nicky Welton, Mike Sweeting, Dawn Lee, Ash Bullement, Nick Latimer, Ed Wilson, Gianluca Baio and Howard Thom for discussions and encouragement, and to Daniel Gallacher regarding the package name.

Software availability and requirements

- Project name: `survextrap`: an R package for survival extrapolation with a flexible parametric model and external data
- Project home page: `https://chjackson.github.io/survextrap`
- Operating system(s): Windows, MacOS and Linux
- Programming language: R and Stan
- Other requirements: R and various R packages, installed automatically
- License: GNU GPL (≥ 3)
- No restriction to use by non-academics

Footnote A Creative Commons licence will be applied to any accepted version of this manuscript.

Author details

MRC Biostatistics Unit, University of Cambridge, Cambridge, UK.

References

1. Bagust A, Beale S. Survival analysis and extrapolation modeling of time-to-event clinical trial data for economic evaluation: an alternative approach. Medical Decision Making 2014;34(3):343–351.
2. Baio G. survHE: survival analysis for health economic evaluation and cost-effectiveness modeling. Journal of Statistical Software 2020;95:1–47.
3. Benaglia T, Jackson CH, Sharples LD. Survival extrapolation in the presence of cause specific hazards. Statistics in Medicine 2015;34(5):796–811.
4. Boag JW. Maximum likelihood estimates of the proportion of patients cured by cancer therapy. Journal of the Royal Statistical Society Series B (Methodological) 1949;11(1):15–53.
5. Bonner JA, Harari PM, Giralt J, Azarnia N, Shin DM, Cohen RB, et al. Radiotherapy plus cetuximab for squamous-cell carcinoma of the head and neck. New England Journal of Medicine 2006;354(6):567–578.
6. Briggs AH, Weinstein MC, Fenwick EAL, Karnon J, Sculpher MJ, Paltiel AD, et al. Model parameter estimation and uncertainty: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force-6. Value in Health 2012;15(6):835–842.
7. Brilleman SL, Elci EM, Novik JB, Wolfe R. Bayesian survival analysis using the rstanarm R package. arXiv preprint arXiv:200209633 2020;.

8. Bullement A, Stevenson MD, Baio G, Shields GE, Latimer NR. A systematic review of methods to incorporate external evidence into trial-based survival extrapolations for health technology assessment. Medical Decision Making 2023;online ahead of print: DOI 10.1177/0272989X231168618.
9. Chaudhary M, Edmondson-Jones M, Baio G, Mackay E, Penrod J, Sharpe D, et al. Use of advanced flexible modeling approaches for survival extrapolation from early follow-up data in two nivolumab trials in advanced NSCLC with extended follow-up. Medical Decision Making 2022;Online ahead of print: DOI 10.1177/0272989X221132257.
10. Che Z, Green N, Baio G. Blended survival curves: a new approach to extrapolation for time-to-event outcomes from clinical trials in health technology assessment. Medical Decision Making 2023;43(3):299–310.
11. Cooney P, White A. Direct incorporation of expert opinion into parametric survival models to inform survival extrapolation. Medical Decision Making 2023;43:325–336.
12. Cope S, Ayers D, Zhang J, Batt K, Jansen JP. Integrating expert opinion with clinical trial data to extrapolate long-term survival: a case study of CAR-T therapy for children and young adults with relapsed or refractory acute lymphoblastic leukemia. BMC Medical Research Methodology 2019;19(1):1–11.
13. Demiris N, Lunn D, Sharples LD. Survival extrapolation using the poly-Weibull model. Statistical Methods in Medical Research 2011;.
14. Federico Paly V, Kurt M, Zhang L, Butler MO, Michielin O, Amadi A, et al. Heterogeneity in survival with immune checkpoint inhibitors and its implications for survival extrapolations: a case study in advanced melanoma. MDM Policy & Practice 2022;7(1):23814683221089659.
15. Gelman A, Hill J, Vehtari A. Regression and Other Stories. Cambridge University Press; 2020.
16. Guyot P, Ades A, Ouwens MJ, Welton NJ. Enhanced secondary analysis of survival data: reconstructing the data from published Kaplan-Meier survival curves. BMC Medical Research Methodology 2012;12:1–13.
17. Guyot P, Ades AE, Beasley M, Lueza B, Pignon JP, Welton NJ. Extrapolation of survival curves from cancer trials using external information. Medical Decision Making 2017;37(4):353–366.
18. Heinze G, Boulesteix AL, Kammer M, Morris TP, White IR. Phases of methodological research in biostatistics — building the evidence base for new methods. arXiv preprint arXiv:220913358 2022;.
19. Ieva F, Jackson CH, Sharples LD. Multi-state modelling of repeated hospitalisation and death in patients with heart failure: the use of large administrative databases in clinical epidemiology. Statistical Methods in Medical Research 2017;26(3):1350–1372.
20. Jackson C, Stevens J, Ren S, Latimer N, Bojke L, Manca A, et al. Extrapolating survival from randomized trials using external data: a review of methods. Medical Decision Making 2017;37(4):377–390.
21. Jackson CH. flexsurv: a platform for parametric survival modeling in R. Journal of Statistical Software 2016;70.
22. Jackson CH, Tom BD, Kirwan PD, Mandal S, Seaman SR, Kunzmann K, et al. A comparison of two frameworks for multi-state modelling, applied to outcomes after hospital admissions with COVID-19. Statistical Methods in Medical Research 2022;31(9):1656–1674.
23. Jansen JP. Network meta-analysis of survival data with fractional polynomials. BMC Medical Research Methodology 2011;11(1):1–14.
24. Jansen JP, Incerti D, Trikalinos TA. Multi-state network meta-analysis of cause-specific survival data. medRxiv 2020;p. 2020–11.
25. Kearns B, Stevenson MD, Triantafyllopoulos K, Manca A. Comparing current and emerging practice models for the extrapolation of survival data: a simulation study and case-study. BMC Medical Research Methodology 2021;21(1):1–11.
26. Król A, Mauguen A, Mazroui Y, Laurent A, Michiels S, Rondeau V. Tutorial in joint modeling and prediction: a statistical software for correlated longitudinal outcomes, recurrent events and a terminal event. Journal of Statistical Software 2017;81(3):1–52.
27. Lambert PC, Royston P. Further development of flexible parametric models for survival analysis. The Stata Journal 2009;9(2):265–290.
28. Latimer NR, Adler AI. Extrapolation beyond the end of trials to estimate long term survival and cost effectiveness. BMJ Medicine 2022;1(1).
29. National Institute for Health and Care Excellence. Guide to the methods of technology appraisal. London: National Institute for Health and Care Excellence; 2013.

30. Nelson CP, Lambert PC, Squire IB, Jones DR. Flexible parametric models for relative survival, with application in coronary heart disease. Statistics in Medicine 2007;26(30):5486–5498.
31. van Oostrum I, Ouwens M, Remiro-Azócar A, Baio G, Postma MJ, Buskens E, et al. Comparison of parametric survival extrapolation approaches incorporating general population mortality for adequate health technology assessment of new oncology drugs. Value in Health 2021;24(9):1294–1301.
32. Ramsay JO. Monotone regression splines in action. Statistical Science 1988;p. 425–441.
33. 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. Statistics in Medicine 2002;21(15):2175–2197.
34. Rutherford M, Lambert P, Sweeting M, Pennington R, Crowther MJ, Abrams KR, et al. NICE DSU Technical Support Document 21: Flexible Methods for Survival Analysis. Decision Support Unit, ScHARR, University of Sheffield 2020;.
35. Spiegelhalter D, Abrams K, Myles J. Bayesian Approaches to Clinical Trials and Health-Care Evaluation. John Wiley and Sons, Chichester, UK; 2004.
36. Spiegelhalter DJ, Best NG, Carlin BP, Van Der Linde A. Bayesian measures of model complexity and fit. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 2002;64(4):583–639.
37. Stan Development Team. Stan Modeling Language Users Guide and Reference Manual; 2022, `https://mc-stan.org`.
38. Stan Development Team, RStan: the R interface to Stan; 2023. `https://mc-stan.org/`, r package version 2.21.8.
39. Tai TA, Latimer NR, Benedict Á, Kiss Z, Nikolaou A. Prevalence of immature survival data for anti-cancer drugs presented to the National Institute for Health and Care Excellence and impact on decision making. Value in Health 2021;24(4):505–512.
40. Vehtari A, Gabry J, Magnusson M, Yao Y, Bürkner PC, Paananen T, et al., loo: Efficient leave-one-out cross-validation and WAIC for Bayesian models; 2020. `https://mc-stan.org/loo/`. R package version 2.4.1.
41. Vehtari A, Gelman A, Gabry J. Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. Statistics and Computing 2017;27:1413–1432.
42. Vervaart M, Aas E, Claxton KP, Strong M, Welton NJ, Wisløff T, et al. General purpose methods for simulating survival data for expected value of sample information calculations. Medical Decision Making 2023;p. 0272989X231162069.
43. Vervaart M, Strong M, Claxton KP, Welton NJ, Wisløff T, Aas E. An efficient method for computing expected value of sample information for survival data from an ongoing trial. Medical Decision Making 2022;42(5):612–625.
44. Vickers A. An evaluation of survival curve extrapolation techniques using long-term observational cancer data. Medical Decision Making 2019;39(8):926–938.
45. Wickham H. Tidy data. Journal of Statistical Software 2014;59(10):1–23.
46. Wood SN. Generalized Additive Models: an Introduction with R. 2nd ed. CRC; 2017.