Checkmate 067 study Bayesian mixture cure model analysis

N Green, UCL

14/10/2020

Background

Immuno-oncologic (IO) studies for melanoma therapies, such as *ipilimumab* (`ipi`), *nivolumab* (`nivo`), and the *nivolumab* with *ipilimumab* (`nivo + ipi`) combination, have indicated that survival curves "plateau" (a considerable proportion of patients are "long-term survivors"). Cure models are a special type of survival analysis where this "cure fraction" (the underlying proportion of responders to treatment/long-term survivors) is accounted for. Cure models estimate the cure fraction, in addition to a parametric survival function for patients that are not cured. The mortality risk in the cured patients is informed by a background mortality rate. The population that is not cured is subject both to background mortality and to additional mortality from their cancer, estimated using a parametric survival model.

A mixture cure model (MCM) (Amico and Van Keilegom (2018)) is a type of cure model where survival is modelled as a mixture of two groups of patients: those who are cured and those who are not (and who therefore remain at risk). The survival for a population with a cure fraction can be written as follows:

$$S(t,x) = S^*(t,x)[\pi(x) + (1-\pi(x))S_u(t,x)], \qquad (*)$$

where $S(t,x)$ denotes the survival at time t, $S^*(t,x)$ denotes the background mortality at time t conditional on covariates x, $\pi(x)$ denotes the probability of being cured conditional on covariates x, and $S_u(t,x)$ denotes the event (progression or mortality) due to cancer at time t conditional on covariates x. For PFS, the survival is composed of either progressing to a disease state or death.

Aims

The aims of the the analysis in this document are as follows:

- Demonstrate the application of a Bayesian mixture cure model using the Checkmate 067 study dataset and the Exponential distribution for event times.
- Produce analogous results to those created previously for the frequentist approach.
- Extend the Bayesian model to incorporate additional structure, including the joint bivariate modelling of OS and PFS event times.

This analysis has been carried-out using the Stan inference engine (Carpenter et al. (2017)) called from R on a Windows PC. The packaged code can be downloaded from GitHub at https://github.com/StatisticsHealthEconomics/rstanbmcm. See the *How to use rstanbmcm* vignette for an introduction to how to use the package.

Model

Let T_i be the non-negative random variable denoting the survival time of patient i with covariate vector x_i.

In the simplest case we can assume that the cure fraction is the same for the whole population i.e. π is fixed. Further, we can assume the π models the relationship between x_i and the probability of being cured. E.g. using a logistic-linear model

$$\pi(x_i|\beta) = 1/[1 + \exp(-x_i^T\beta)].$$

The likelihood of the standard survival is

$$L = \prod_i S(t_i|x_i)h(t_i|x_i)^{\delta_i}$$

Log-likelihood is therefore

$$l = \sum_i \log(S(t_i|x_i)) + \delta_i \log(h(t_i|x_i))$$

Plugging this directly into the mixture cure equation in (*) gives

$$l(\pi|\delta, x) = \sum_i \log(S^*(t_i|x_i)h^*(t_i|x_i)^{\delta_i}[\pi(x) + (1 - \pi(x))S_u(t_i|x_i)h_u(t_i|x_i)^{\delta_i}])$$

We will assume that the cured component is the exponential survival model. The non-cured component can be thought of in similar terms to the cumulative incidence function. That is, the probability of an event is the combined probability of surviving both events (e.g. for OS, all-cause and cancer mortality) and then experiencing either i.e. dropping the S dependencies for brevity

$$S^*S_u(h^*)^\delta + S^*S_u(h_u)^\delta = S^*S_u(h^* + h_u)^\delta \qquad (**)$$

Bayesian formulation

In a Bayesian approach to modelling, all quantities that are subject to uncertainty are modelled using probability distributions. This applies to observed data (e.g. time to PFS for a given individual), that are subject to sampling variability, as well as to unobservable parameters (e.g. the coefficient quantifying the impact of age or sex over the average survival curve). In this latter case, probability distributions are used to model the epistemic uncertainty (e.g. the fact that we do not know for certain what the "true" underlying value of the model parameter is). In addition, we may model as yet unobserved (but potentially observable) quantities using a suitable probability model. For example, we could consider the extrapolated part of the survival curve as subject to uncertainty due to the current sampling process giving rise to the data that are actually observed, as well as the uncertainty on the underlying data generating process.

We can mix different sources of evidence to form our "prior" distributions, which are used to describe the state of science on the model parameters. These are then combined with any observed data to form an updated level of knowledge. This process is particularly relevant in the case at hand, when data can only inform about limited aspects of the overall underlying reality. For this reason, it is important to a) include information/evidence available in the form of external data and/or expert opinion; b) extract the most information possible from the available data (e.g. by formally trying to model the correlation between the PFS and the OS data to borrow strength from the more mature set of observations).

A built-in advantage of the Bayesian procedure is that uncertainty is directly and formally propagated to an economic model; the main output from the statistical analysis (the extrapolated survival curve) are produced by default as based on a full posterior distribution. From this, we can easily derive a "base case" (e.g. taking the mean value) but without the need for further tools (such as bootstrap) we already have a full characterisation of the underlying uncertainty that can be used in the process of probabilistic sensitivity analysis. We can moreover add information in the priors to ensure that the extrapolation beyond the observed data is realistic and consistent with the clinical expertise (e.g. by "anchoring" the extrapolated survival curve to be probabilistically below the curves for the healthy population, or by ensuring that OS behaves in a way to respect some agreed level of similarity, or correlation, to PFS).

The general form of the posterior distribution can be written as follows.

$$p(\pi, \beta^u, \beta^*|\delta, x) \propto L(\pi, \beta^u, \beta^*|\delta, x) f(\pi) g_2(\beta^u) g_3(\beta^*)$$

assuming that the cure fraction is independent of the covariates.

Cure fraction

There are two obvious ways to represent the uncertainty about the cure fraction in the model.

The first is to specify the cure fraction directly using a $Beta(a_{cf}, b_{cf})$ prior, most uninformative as a uniform $Beta(1, 1)$. The parameters can be obtained via transformation of mean and sd to allow a more natural scale for elicitation.

Alternatively, we may specify the uncertainty on the real line with a Normal distribution and then transform to the probability scale.

A further consideration is how to represent the cure fraction so to share information between the OS and PFS data. We will investigate 3 alternatives.

- *Pooled*: Assume that the cure fraction is the same for OS and PFS i.e. $cf_{os} = cf_{os} = cf$ where
$$logit(cf) \sim N(\mu_{cf}, \sigma^2_{cf}),$$
- *Separate*: Model each independently.
$$logit(cf_{os}) \sim N(\mu_{cfos}, \sigma^2_{cfos}), \;\; logit(cf_{pfs}) \sim N(\mu_{cfpfs}, \sigma^2_{cfpfs})$$
- *Hierarchical*: Assume exchangeability between OS and PFS
$$cf \sim N(\mu_{cf}, \sigma^2_{cf}), \;\; logit(cf_{os}) \sim N(cf, \sigma^2_{cfos}), \;\; logit(cf_{pfs}) \sim N(cf, \sigma^2_{cfpfs})$$

Exponentially distributed event times

First, consider the simplest case where both the background and cancer times to event follow exponential distributions.

Define $f(t)$ density, $S(t)$ survival and $h(t)$ hazard functions.

$$f(t) = \lambda \exp(-\lambda t), \;\; S(t) = \exp(-\lambda t), \;\; h(t) = \lambda$$

Which gives the likelihood

$$l(\pi|\delta, x) = \sum_i \log(\exp(-\lambda^* t)\lambda^{*\delta_i}[\pi(x) + (1 - \pi(x)) \exp(-\lambda_u t)\lambda_u^{\delta_i}])$$

Substituting $S(t)$ and $h(t)$ into (**)

$$f_u^* = e^{-\lambda^* t} e^{-\lambda_u t} (\lambda^* + \lambda_u)^\delta \quad i.e. \text{for no censoring } \; T \sim Exp(\lambda^* + \lambda_u)$$

Background survival

The frequentist analysis used the World Health Organization (WHO) life tables by country for the latest year available of 2016 (WHO (2020)) to inform the background mortality rate (baseline hazard). These baseline hazards are the expected mortality rate for each patient at the age at which they experience the event. The mortality data are age- and gender adjusted, thus providing a granular account of the different patient profiles in the trial. The WHO reports conditional probabilities of death in 5-year intervals until age 85. A constant annual mortality rate is reported for individuals over 85. They assumed that the maximum age is 100 years.

In our analysis there are alternative ways in which we could model the background mortality.

Use WHO hazard point estimates as known We could consider the WHO estimates to provide sufficiently accurate estimates given the sample size and so incorporating uncertainty is not necessary.

Survival distribution informed by the WHO data However, we also want the developed model to be able to be applied to other data sets which may be smaller or noisy. Also the mortality rate for the cured study population may not be the same as the general population. Sensible prior parameter values can be taken for the life table hazard curve. After infancy the log-hazard is approximately linear and so intercept and slope estimates are simple to obtain.

Model checking

Before using the real data we developed the models using simulated data sets. In this way we know the true underlying data generating process. It is trivial to simulate random variables from the given distributions in R using the in-built functions. However, times from the 'product' distribution used for the non-cured combined background and cancer event times need to be generated from a bespoke function.

We have shown above that the density can be thought of as the probability of surviving both events up to time t and then experiencing either at time t. This is equivalent to the earliest event time being after time t i.e.

$$P(\min\{X, Y\} > t) = P(X > t, Y > t) = P(X > t)P(Y > t) = S_X S_Y$$

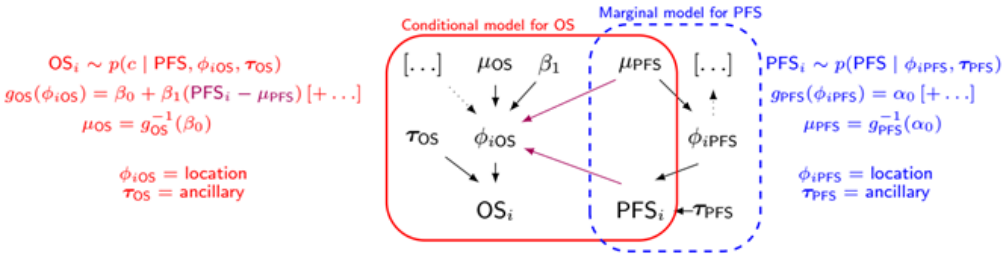
Thus, we simulate all latent event times and select the smallest as the observed time. In the simplest case, this is a well-known method for simulating jump times in a Markov model. More generally, we can think of the exponential-weibull model as a version of a poly-hazard model.

Jointly distributed event times model

One effective way of modelling joint (bivariate) distributions is to factorise them into a marginal and a conditional distribution (which holds as a fundamental rule of probability). In general terms, we can then write $p(x, y) = p(x)p(y|x)$. In the context of our model, we can use this intuition to model the joint distribution of the PFS and OS observed times (in terms of their survival curves) as:

$$S(t_{OS}, t_{PFS}) = P\ (T_{OS} \geq t_{OS}, T_{PFS} \geq t_{PFS}) = S_{PFS}(t_{PFS}) S_{OS|PFS}(t_{OS} \,|\, T_{PFS} = t_{PFS}).$$

The structure above implies that essentially we first create a marginal generalised linear regression to model the survival curve for the PFS data (as a function of relevant covariates); the second module of the model implies another generalised linear regression for the OS data, where the observed PFS data act as a covariate (in addition to other relevant predictors, which may or may not be the same used for the PFS model). Alternative specifications are possible (for instance, the generalised linear model can be applied on the scale of the hazard function, if more appropriate). This modelling approach can be visualised in the graph below.



If we factorise into a marginal and conditional components to model the underlying bivariate distribution can be written generally as

$$OS_i \sim p(c|PFS, \phi_{iOS}, \tau_{OS})$$

$$g_{OS}(\phi_{iOS}) = \beta_0 + \beta_1(PFS_i - \mu_{PFS})[+ ...]$$
$$\mu_{OS} = g_{OS}^{-1}(\beta_0)$$
$$g_{PFS}(\phi_{iPFS}) = \alpha_0[+ ...]$$
$$PFS_i \sim p(PFS|\phi_{iPFS}, \tau_{PFS})$$
$$\mu_{PFS} = g_{PFS}^{-1}(\alpha_0)$$

The combined log-likelihood is

$$l = l_{OS} + l_{PFS}$$

For the case with exponential OS times and exponential PFS times with centred age this gives the following.

$$t_{iOS} \sim Exp(\phi_{iOS})$$
$$\log(\phi_{iOS}) = \beta_0 + \beta_1(t_{iPFS} - \bar{t}_{PFS}) + \beta_2 age_{iPFS}$$
$$\mu_{OS} = \exp(\beta_0)$$
$$\log(\phi_{iPFS}) = \alpha_0 + \alpha_1 age_{iOS}$$
$$t_{iPFS} \sim Exp(\phi_{iPFS})$$
$$\mu_{PFS} = \exp(\alpha_0)$$
$$\bar{t}_{PFS} = 1/\mu_{PFS}$$

Results

We fit the exponential hazard model to the study data and produced the posterior survival curves below.

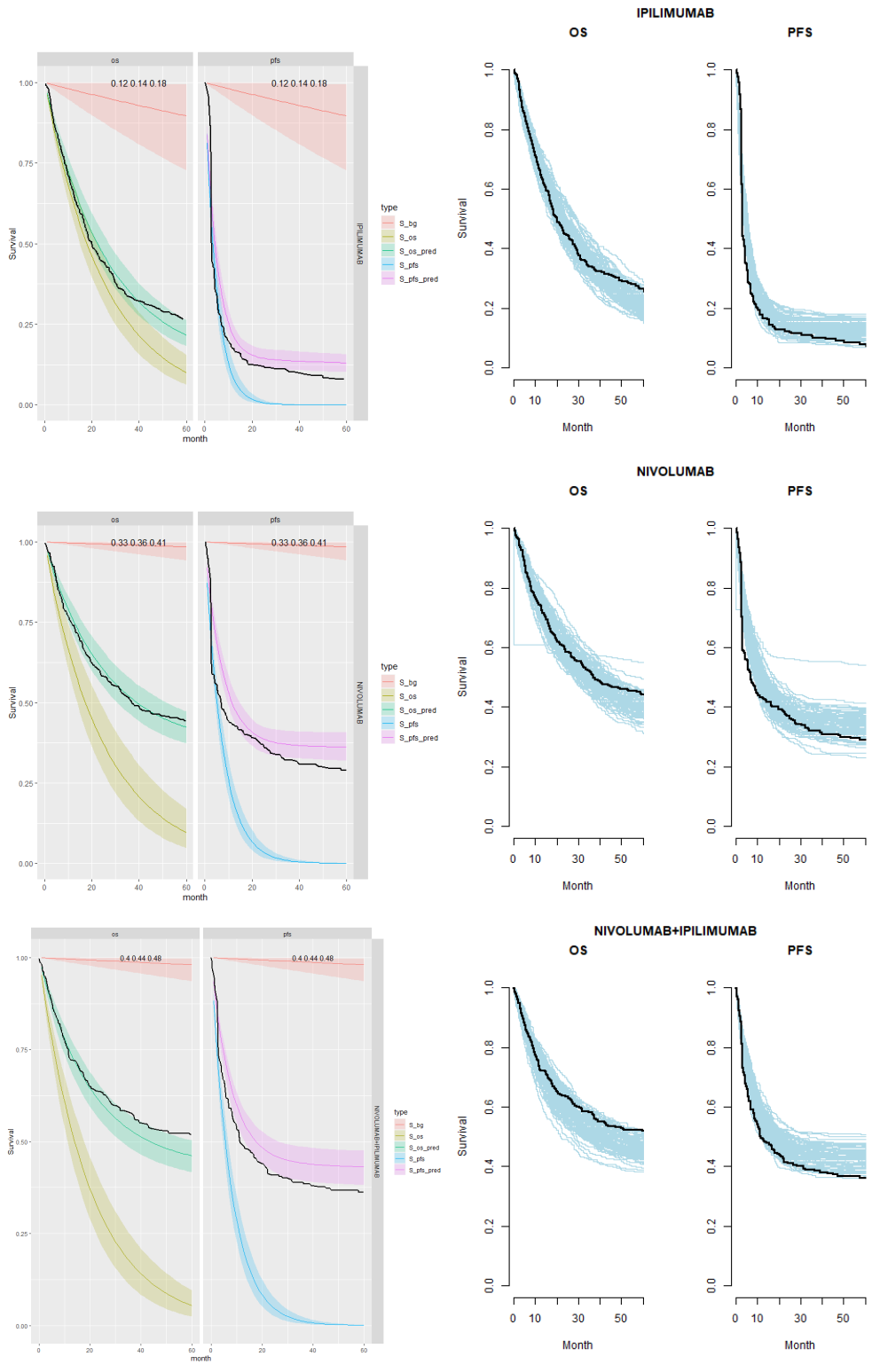
For each model and treatment we produce two figures:

1. The expected survival curves with 95% Credible Intervals (CrI). The OS curves are to the left-hand side and PFS curves to the right-hand side. Background mortality (i.e. cured patients) is indicated by the red line. Non-cured patients survival curves are shown in dark green and blue for OS and PFS respectively. Light green and magenta are the total sample. The black line is the Kaplan-Meier curve for the observed data. Note that these plots are for an average individual, e.g. at average age, and so we would not expect them to perfectly match the sample data Kaplan-Meier.
2. Kaplan-Meier curves for 50 simulated trials using the posterior prediction distribution for time to event. This is a good model checking plot, indicating if we can replicate data similar to the observed data using the fitted model.

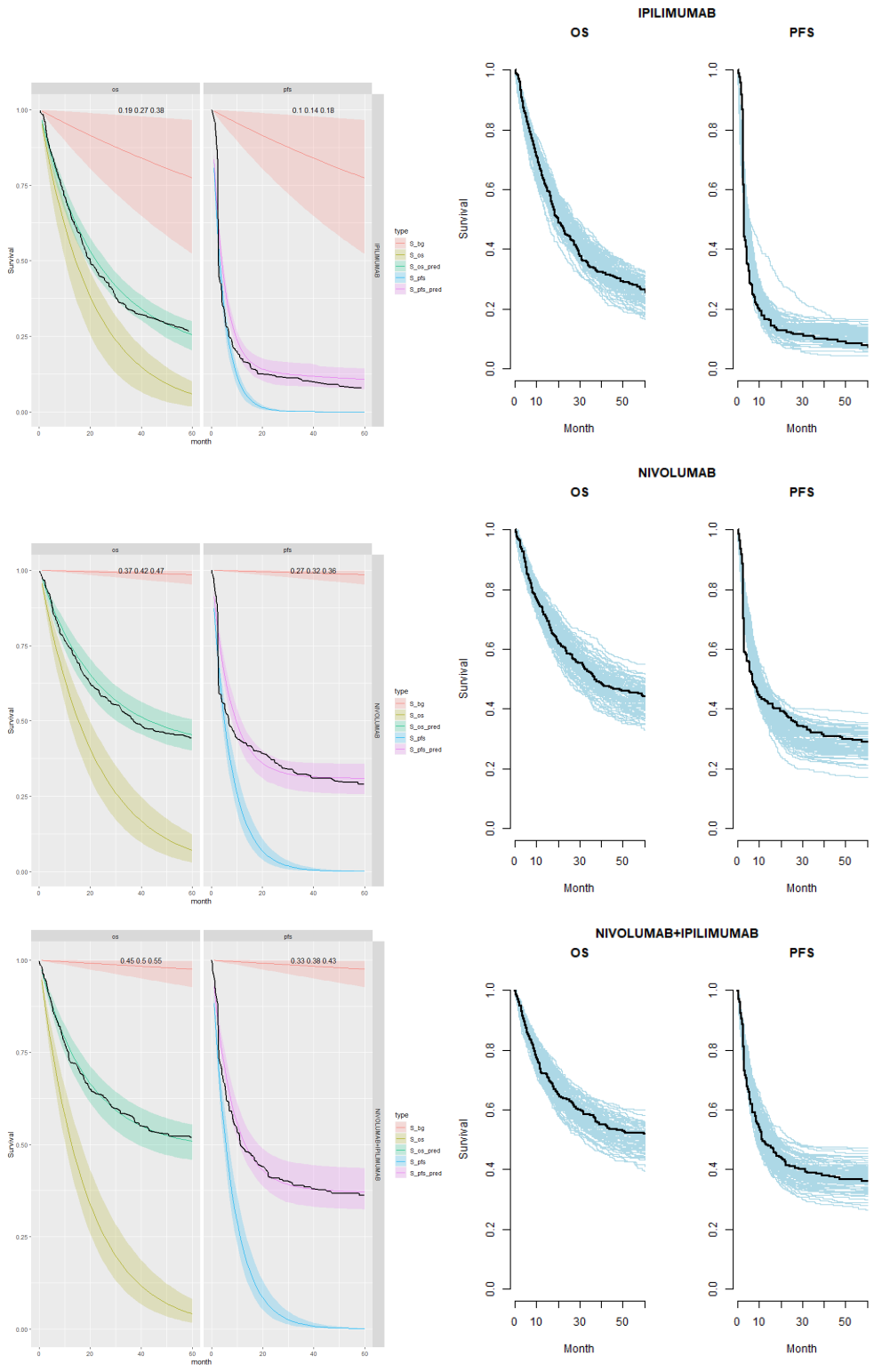
For the Checkmate dataset we found that the hierarchical model gave equivalent results to the separate cure fraction model and so for speed of computation and simplicity we only show the separate cure fraction results here.

Independent PFS and OS event times

Pooled cure fraction This is the most restrictive model and so as we would expect it gives the worse results. The OS is not bas but the PFS CrIs fail to contain the observed data.



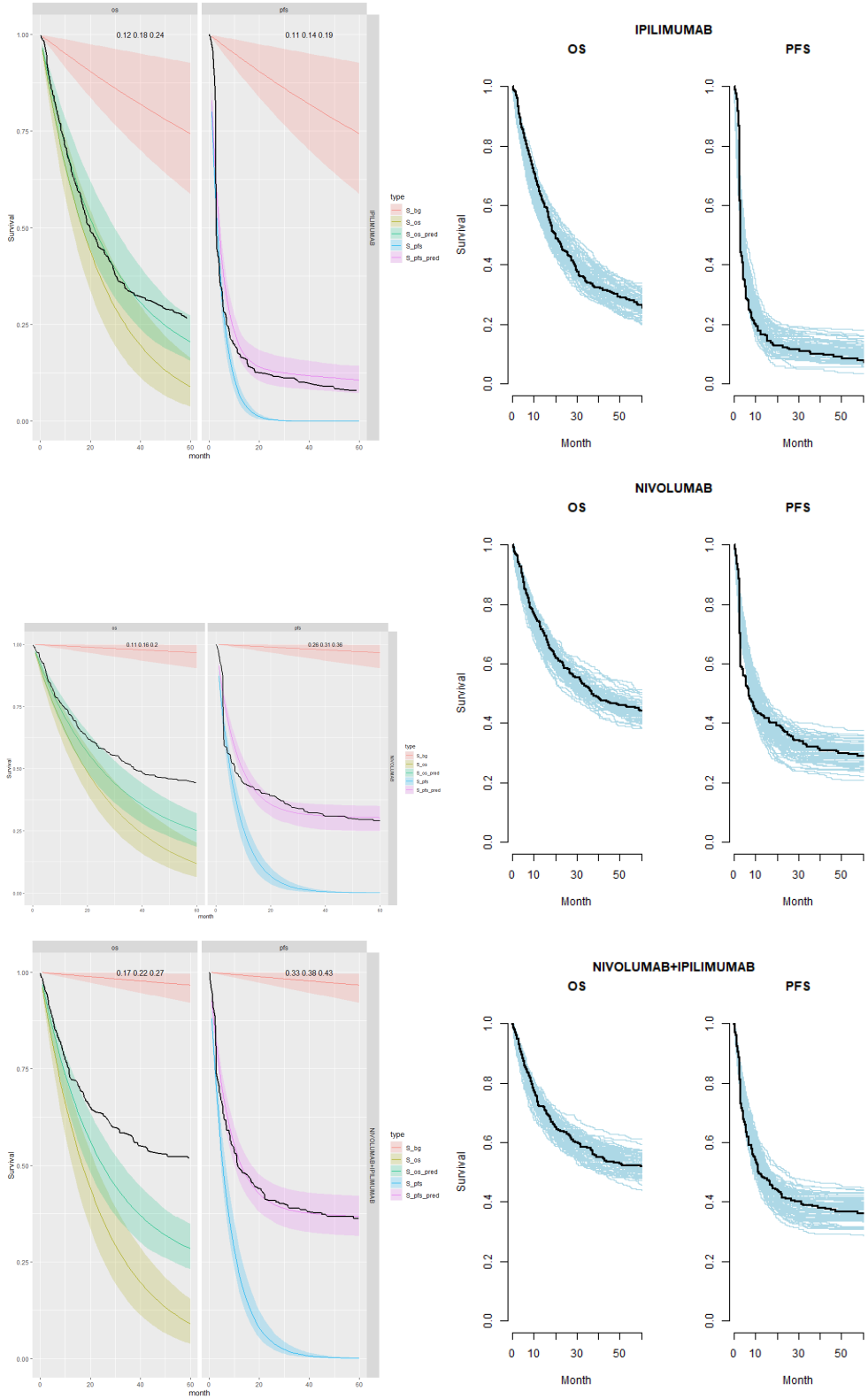
Separate cure fraction



Jointly distributed PFS and OS event times

We see that the expected OS survival curves are biased; over-estimating the rate. This is not the case for the *ipilimumab*. This is due to the additional joint component in the OS linear regression. This may be due to censoring since *ipilimumab* has the smallest amount. This is an area for research.

However, the model does fit well taking in to account the case-mix of the study population. The posterior prediction plots show that the Kaplan-Meier for the observed data lies within the simulated curves.



The table below summarises the cure fraction posterior distribution for each scenario.

Event times	Cure fraction	Treatment	cf_{OS} (CrI)	cf_{PFS} (CrI)
Independent	Pooled	IPI	0.14 (0.12, 0.18)	0.14 (0.12, 0.18)
Independent	Pooled	NIVO	0.36 (0.33, 0.41)	0.36 (0.33, 0.41)
Independent	Pooled	NIVO+IPI	0.44 (0.4, 0.48)	0.44 (0.4, 0.48)
Independent	Separate	IPI	0.27 (0.19, 0.38)	0.14 (0.1, 0.18)
Independent	Separate	NIVO	0.42 (0.37, 0.47)	0.32 (0.27, 0.36)
Independent	Separate	NIVO+IPI	0.5 (0.45, 0.55)	0.38 (0.33, 0.43)
Joint	Separate	IPI	0.18 (0.12, 0.24)	0.14 (0.11, 0.19)
Joint	Separate	NIVO	0.16 (0.1, 0.2)	0.31 (0.26, 0.36)
Joint	Separate	NIVO+IPI	0.22 (0.17, 0.27)	0.38 (0.33, 0.43)

Future work

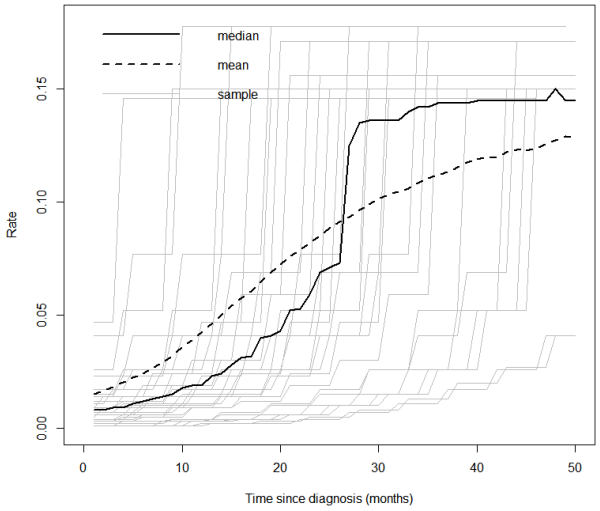
Expand available distributions Having shown the application of these methods to this problem we will extend the tool kit to include other standard parametric distributions are tested:

- Weibull
- Gompertz
- Log-normal
- Log-logistic
- Generalised gamma.

Sensitivity analysis Of course, when the data offer only limited amount of information, the assumptions in the prior distribution possibly exert much influence on the results - and crucially on the decision model output. We will conduct extensive sensitivity analysis and will justify assumptions in all aspects of the modelling strategy by assessing the meaning of the various distributional assumptions visually and formally.

Background survival An alternative non-parametric approach, as used in Demiris and Sharples (2006), is to use a *Gamma process* to define gamma distributions at each time. A variance parameter determines the influence between times.

Average values derived from the life tables are used in the Gamma process. These are age-sex-country standardised. The below plot shows the mean, median and a sample of hazard curves for the checkmate data set. The underlying hazard curve for 0-100 year olds is shifted left depending on the starting age of an individual in the cohort.



Additional model structure The two regression models can be extended to more complex structures, for instance by including a mixture model in one or both (and eventually by including some further correlation structure in the mixing parameters).

For this analysis it appeared unnecessary but we may generalise the model so that the cure fraction is dependent on covariates.

$$p(\beta^u, \beta^*, \beta^{cf}|\delta, x) \propto L(\beta^u, \beta^*, \beta^{cf}|\delta, x) g_1(\beta^{cf}) g_2(\beta^u) g_3(\beta^*)$$

References

Amico, Maïlis, and Ingrid Van Keilegom. 2018. "Cure Models in Survival Analysis." *Annual Review of Statistics and Its Application* 5: 311–42. https://doi.org/10.1146/annurev-statistics-031017-100101.

Carpenter, Bob, Andrew Gelman, Matthew D Hoffman, Daniel Lee, Ben Goodrich, Michael Betancourt, Marcus Brubaker, Jiqiang Guo, Peter Li, and Allen Riddell. 2017. "Stan: A Probabilistic Programming Language." *Journal of Statistical Software* 76 (1).

Demiris, Nikolaos, and Linda D. Sharples. 2006. "Bayesian evidence synthesis to extrapolate survival estimates in cost-effectiveness studies." *Statistics in Medicine* 25 (11): 1960–75. https://doi.org/10.1002/sim.2366.

WHO. 2020. "Life tables by country." http://apps.who.int/gho/data/?theme=main&vid=61160.