Parametric multi-state survival models: flexible modelling allowing transition-specific distributions with application to estimating clinically useful measures of effect differences

Michael J. Crowther[1,2][*][†] and Paul C. Lambert[1,2]

Multi-state models are increasingly being used to model complex disease profiles. By modelling transitions between disease states, accounting for competing events at each transition, we can gain a much richer understanding of patient trajectories and how risk factors impact over the entire disease pathway. In this article we concentrate on parametric multi-state models, both Markov and semi-Markov, and develop a flexible framework where each transition can be specified by a variety of parametric models including exponential, Weibull, Gompertz, Royston-Parmar proportional hazards models or log-logistic, log-normal, generalised gamma accelerated failure time models, possibly sharing parameters across transitions. We also extend the framework to allow time-dependent effects. We then use an efficient and generalisable simulation method to calculate transition probabilities from any fitted multi-state model, and show how it facilitates the simple calculation of clinically useful measures, such as expected length of stay in each state, and differences and ratios of proportion within each state as a function of time, for specific covariate patterns. We illustrate our methods using a dataset of patients with primary breast cancer. User friendly Stata software is provided. Copyright © 0000 John Wiley & Sons, Ltd.

Keywords: survival analysis; multi-state models; competing risks

[1] *University of Leicester, Department of Health Sciences, Adrian Building, University Road, Leicester LE1 7RH, UK.*
[2] *Karolinska Institutet, Department of Medical Epidemiology and Biostatistics, Box 281, S-171 77 Stockholm, Sweden.*
[*] *Correspondence to: Michael J. Crowther. University of Leicester, Department of Health Sciences, Adrian Building, University Road, Leicester LE1 7RH, UK.*
[†] *E-mail: michael.crowther@le.ac.uk*

1. Introduction

Multi-state models allow rich insights into complex disease pathways, where a patient may experience many non-fatal/intermediate events, and we wish to investigate covariate effects for each specific transition between two states, not just for example, on the first event, or a terminal event [1].

Methodological development has generally focused on semi-parametric approaches, utilising the Cox model as the fundamental framework of the multi-state model. As with standard survival analysis, use of the Cox model avoids making assumptions for the underlying shape of the baseline hazard function, and indeed thanks to the widely used `mstate` package in R [2], the methods are highly accessible to researchers. We argue that the benefits of a fully parametric approach are less well known, and indeed should be used more extensively [3]. With the drive towards personalised medicine, and expanded availability of large registry-based data sources, including data-linkage, there are substantial opportunities to gain greater understanding of disease processes, and how they change over time. Through a parametric approach, there are a number of distinct advantages, such as; prediction, extrapolation, quantification (i.e. absolute and relative measures of risk), ease of modelling time-dependent effects (non-proportional hazards) in comparison to semi-parametric approaches, and a greater understanding of how risk changes and evolves over time [4]. A particularly important area where parametric models must be employed is within a health economic framework, as often it is required to extrapolate survival to a lifetime horizon [5]. Of course, extrapolation is fraught with dangers, and arguably should only be attempted in the presence of appropriate external data [6].

Previous parametric multi-state models can be considered limited, in the fact that often the same distribution must be assumed for all transitions, such as in the extensive `flexsurv` package in R [7]. For example, Omar et al. (1995) [8] developed an illness-death model assuming Weibull transitions. Piecewise exponential multi-state models have been implemented in R by Jackson (2011) [9], with a focus on panel data. This allows the convenience of the exponential distribution and indeed can capture complex shapes; however, assuming a discontinuous baseline hazard function is often not biologically plausible, which has motivated other approaches [10]. Titman (2011) [10] proposed smooth hazard models for non-homogeneous Markov models through the use of quadratic B-splines to model the transition specific baseline hazard functions, and employed a differential equation solver to calculate transition probabilities. Król and Saint-Pierre recently developed the SemiMarkov package in R which allows transition specific distributions for semi-Markov models, but the only distributions available are the exponential, Weibull and exponentiated-Weibull, which can still be considered too simplistic for many practical situations [11]. Blaser et al. (2015) developed a flexible approach, implemented in the `gems` package in R, allowing transition-specific models; however, used a piecewise approximation to simulate event times from the fitted models. Others have employed numerical integration to calculate transition probabilities, but this is generally limited to illness-death models, as anything more complex than a two-dimensional numerical integration can become highly computationally intensive and loss of accuracy can occur [12, 13]. Arguably, transition-specific distributions in a multi-state setting is relatively simple, as it's fundamentally a matter of data setup and fitting models using standard software [1]. The challenge lies in calculating predictions from a given multi-state model.

In this article we concentrate on Markov uni-directional models, and assume transition times are observed exactly. We develop a framework allowing the specification of a model for each transition, coming from possibly different parametric distributions or functional forms, also allowing for parameters to be shared across transitions (regardless of the transition specific distribution chosen). In many situations there will be sparse data for specific transitions and we may wish to assume simple distributions such as the exponential or Weibull, and where we have more information for a transition, we can apply a more flexible model such as the generalised gamma [14] or a spline based specification [15]. We then incorporate time-dependent effects into any transition model. We adopt the simulation approach to calculate transition probabilities and other quantities of interest [16, 17, 18, 7], using a general survival simulation

algorithm which can be readily applied to any combination of transition specific distributions to simulate from the multi-state model [19]. Finally, we show the flexibility of the simulation approach to obtain clinically useful measures, such as total length of stay [20], and differences and ratios of proportions within each state for specific covariate patterns, and accompanying confidence intervals.

The article is organised as follows. In Section 2, we introduce the multi-state modelling framework for Markov models, briefly describing previous approaches. In Section 3, we develop multi-state models allowing for transition-specific distributions, and the general simulation approach to calculating transition probabilities. In Section 4, we extend multi-state models to incorporate time-dependent effects (non-proportional hazards). In Section 5 we describe other clinically useful measures and contrasts. In Section 6, we apply our methods to a dataset of breast cancer patients who begin in a healthy state and can experience relapse, death, or death after relapse. We conclude the paper in Section 7 with a discussion.

2. Multi-state models

We begin with a brief introduction to multi-state models, describing how to fit commonly used models, before describing our proposed extensions in the next section. Following [16], consider a random process Y(t), t ≥ 0 which takes the values in the finite state space $\mathcal{Z} = \{1, \ldots, Z\}$. We define the history of the process until time s, to be $\mathcal{H}_s = \{Y(u); 0 \leq u \leq s\}$. The transition probabilities can then be defined as,

$$P(Y(t) = b | Y(s) = a, \mathcal{H}_{s-}) \tag{1}$$

where $a, b \in \mathcal{Z}$. This is the probability of being in state b at time t, given that it was in state a at time s and conditional on the past trajectory until time s. A Markov multi-state model makes the following assumption,

$$P(Y(t) = b | Y(s) = a, \mathcal{H}_{s-}) = P(Y(t) = b | Y(s) = a) \tag{2}$$

which implies that the future behaviour of the process is only dependent on the present state. $\mathcal{H}_{s-}$ is the history of the process up to a time just before s. The transition intensity is then defined as,

$$h_{ab}(t) = \lim_{\delta t \to 0} \frac{P(Y(t + \delta t) = b | Y(t) = a)}{\delta t} \tag{3}$$

alternatively, and more conveniently for the remainder of the article, we can describe the model indexing by transition. Therefore, for the kth transition from state a_k to state b_k, we have

$$h_k(t) = \lim_{\delta t \to 0} \frac{P(Y(t + \delta t) = b_k | Y(t) = a_k)}{\delta t} \tag{4}$$

which represents the instantaneous failure rate of moving from state a_k to state b_k. The Markov assumption can be relaxed, to become a semi-Markov model, where the transition probabilities and subsequently the intensities can depend on the time at which earlier states were entered. This is commonly simplified further, by defining the transition intensities to be dependent only on the time spent in the current state, as opposed to time spent in previous states [1]. We discuss the semi-Markov extension in Section 7.

2.1. Estimation

A multi-state model can be specified by a combination of transition-specific survival models. The most convenient way to do this is through the stacked data notation (see the appendix for an example), described in the excellent

tutorial paper by Putter et al. (2007) [1]. A patient has a row of data for each transition that they are at risk for, using start and stop notation. We can therefore specify a model for the kth transition,

$$h_k(t; X) = h_{0k}(t) \exp(X_k \beta_k) \tag{5}$$

where X_k is a vector of time-independent covariates, possibly different for different transitions, with associated log hazard ratios β_k. Once the data is setup, a model can be fitted using any standard survival command, for example allowing transition intensities to be proportional to a reference transition, or stratify by transition, and allow covariate effects to be constrained to be equal across transitions. However, fitting a parametric multi-state model to the stacked data imposes the limitation that all transitions must be modelled using the same distribution. This is a limitation which we address in this paper.

2.2. Calculating transition probabilities

Obtaining fitted/predicted transition intensities is relatively simple as when using a parametric approach we have an analytical expression; however, calculating transition probabilities is generally non-trivial. When using a parametric approach, i.e. modelling the baseline hazard function for each transition (in any form be it separate or stacked data models), generally the calculation of transition probabilities is computationally challenging. Early work concentrated on the use of the exponential distribution, as it provides a computationally convenient solution to the forward Kolmogorov equation, allowing direct calculation of the transition probabilities. A simple extension is to specify piecewise-exponential distributions, partitioning the data at user-defined cut-points and assuming a constant baseline hazard function within each interval. This allows the convenience of the exponential distribution and indeed can capture complex shapes; however, assuming a discontinuous baseline hazard function is often not biologically plausible, and can result in many model parameters, which has motivated other approaches [10]. Titman (2011) [10] proposed smooth hazard Markov models through the use of quadratic B-splines to model the transition specific baseline hazard functions, and employed a differential equation solver to calculate transition probabilities. Some authors have also considered numerical integration [12, 21], but this can be considered limited in its generalisability.

In this article we concentrate on the simulation approach, which can be considered the most convenient and general approach. The benefits of the approach becomes even more apparent in later sections focusing on extended predictions. We begin by simulating a large sample, N (we use a default of 10,000), using the fitted transition intensities, and iterating through the transition matrix until all patients have reached an absorbing state or are censored at a specified maximum follow-up time [16, 18]. This is done by viewing the simulation as a sequence of competing risk scenarios. Following Fiocco et al (2008) [16], we describe the procedure as follows:

Let a be the starting state, entered at time $t_a = s$. We specify a maximum follow-up time to be $t_{\max}$. Repeat the following,

1. Let $\mathcal{B}$ be the set of states that can be reached from state a, and N_a the number of possible next states. If $\mathcal{B} = 0$, stop. Otherwise, let, for $b \in \mathcal{B}$, $h_{ab}(t)$ be the hazard function for transition $a \to b$.
2. Using $h_{ab}(t)$, sample a survival time using the general inversion method described by Crowther and Lambert [19], for each possible transition, denoted by t_{ab}^*.
3. Set $t^* = \min\{t_{a1}^*, \ldots, t_{aN_a}^*, t_{\max}\}$, if $t^* = t_{\max}$, stop.
4. Set $a = c$, where $t^* = t_{ac}$, $c \in \mathcal{B}$ and set $t_a = t^*$.

We simply repeat the above procedure for a large N number of paths through the multistate model, with the same covariate vector, and then at various time points, count the number of paths for which each event occurred, to give us our patient-specific transition probabilities. The appeal of this approach is in its generality, as it's immediately applicable to both Markov and non-Markov models. The associated Stata package can calculate transition probabilities through simulation, after fitting a variety of parametric survival models on the stacked data,

including the exponential, Weibull, Gompertz, log-logistic, log-normal, generalised gamma, and Royston-Parmar models. Simulating from standard distributions is computationally efficient, and even in situations where there is no tractable simulation solution, such as for the Royston-Parmar model, we can employ the general survival simulation algorithm of Crowther and Lambert (2013) [19]. This allows us to incorporate time-dependent effects, or model on alternative scales, for example using an accelerated failure time model.

Furthermore, we can conduct a single large simulation, with a specified maximum follow-up t_{max}, and then for a vector of times between $\mathbf{t} = \{0, \ldots, t_{max}\}$, we can calculate the transition probabilities at each time point using the single generated dataset, because we have each patient's state and transition time history. This provides a highly efficient way of producing graphical displays of transition probabilities, such as those seen in Figure 3, and extended predictions.

2.2.1. Confidence intervals Following the fit of a multi-state model, we obtain our vector of estimated coefficients (be they log hazard ratios or log acceleration factors), $\hat{\beta}_{ab}$, for the $a \to b$ transition. As we are currently assuming that the same underlying distribution is being used, i.e. a single model has been fitted to the stacked data, we have an overall parameter vector, $\mathbf{b}$, and associated variance-covariance matrix, $\mathbf{V}$. To obtain confidence intervals for the predicted transition probabilities, we follow the approach of Fiocco et al. (2008) [16], where we can simply draw from a multivariate normal distribution with mean $\mathbf{b}$, and variance $\mathbf{V}$, calculate the transition probabilities at the specified time points for a given simulated sample of size N, and repeat a total of M times. We can then construct 95% confidence intervals by calculating the variance of our sets of M estimates. Due to the expected computational burden of the nested simulations, Fiocco et al (2008) proposed a method of variance adjustment, which allowed them to reduce M and subsequently apply a variance correction formula to obtain unbiased estimates. In their article this allowed them to use N=10,000 and M=50. However, as we discuss later, our implementation is very efficient and hence in our article we use N=10,000 and M=200 to obtain similar levels of accuracy. The correction formula of Fiocco et al. (2008) will be added to the associated software package.

3. Extending multi-state models to transition-specific distributions

As described in the previous section, we can use any standard survival model fitting command to estimate the transition intensities, once the data is correctly setup. However, this imposes the limitation that all transitions must come from the same distributional model, e.g. all transitions must assume a Weibull baseline function, which could of course be common across transitions, proportional between transitions, or separate (stratified) across transitions. In this section we relax this assumption, to allow each transition to come from potentially different parametric distributions, producing substantial flexibility in model fitting.

3.1. Transition specific distributions

The stacked data format, which is often used when fitting multi-state models, is an extremely convenient way to conduct such analyses. However, it is not strictly necessary when fitting transition specific hazard models. Models can be fitted separately to each transition, which can actually be more computationally efficient, compared to fitting the full model on the stacked data. By fitting models separately to each transition, this crucially allows the use of different distributions for each transition. This is important as often there will be limited data for particular transitions, and hence being able to fit an exponential distribution for example, for a sparse transition, and more complex distributions for other transitions, allows an effective strategy to still make use of all the available data.

Consider a set of available covariates X. We therefore define, for the kth transition, the hazard function at time t is,

$$h_k(t) = h_{0k}(t) \exp(X_k \beta_k) \tag{6}$$

where $h_{0k}(t)$ is the baseline hazard function for the $a \to b$ transition, which we now allow to take any parametric form such that $h_{0k}(t) > 0$. Previously in Equation (5) we assumed the same underlying distribution for all transitions. To maintain flexibility, we have a vector of patient-level covariates included in the $a_k \to b_k$ transition, X_k, where $X_k \in X$. This allows different covariates to be included in different transitions, or indeed the same set. We also have the associated log hazard ratios, β_k. In the associated Stata package, we provide a variety of choices including the exponential, Weibull, Gompertz, log-logistic, log-normal, generalised gamma, and Royston-Parmar model. This allows combinations of proportional hazards models and accelerated failure time models, emphasising the flexibility of this framework.

If no parameters are shared across transitions, then a very complex and flexible multi-state model can be fitted by considering each transition to be a separate survival analysis. In other words, the likelihood is the same when fitting separate models to each transition, compared to fitting a model to the stacked data, if the baseline and covariate effects are completely separate. Alternatively, to maintain the convenience of the stacked data setup, we propose a model fitting routine, which maximises the combined likelihood, across transitions, allowing for different distributions for each transitions. With either separate models, or a model fitted to the stacked data, the challenge is then to calculate transition probabilities, which we describe in Section 3.3.

3.2. Sharing parameters across transitions

One of the benefits to the stacked data approach is the ability to share parameters across transitions. This can be done, for example, when we wish to assume that a covariate's effect is equal across various transitions, or indeed to incorporate proportionality between baseline hazard functions within the same distributional model, across transitions [2]. This may be useful if you have sparse data for a particular transition. In the transition specific multi-state model described in the previous subsection, we can allow the sharing of covariate effects across transitions simply by imposing constraints on relevant parameters during the model fitting.

3.3. Calculating transition probabilities

As in Section 2.2, we adopt the simulation approach, using each transition-specific hazard model to simulate individual survival times as each observation moves through the transition matrix until they reach an absorbing state or are censored at the maximum follow-up time. In the accompanying Stata software, these calculations can be conducted either when completely separate models are fitted to each transition, or if a transition-specific model is fitted to the full stacked data. In the associated Stata software, we use as a default sample size of 10,000.

3.3.1. Confidence intervals To calculate confidence intervals, we can once again draw from a multivariate normal distribution, with mean the estimated parameter vector and variance the associated variance-covariance matrix. If separate models are fitted, then parameters between transition models are independent and we can therefore sample from $\text{MVN}(\mathbf{b}_k, \mathbf{V}_k)$, for each transition (or alternative stack the parameter vectors, and fill in the appropriate covariances with zeroes). If fitting a transition-specific model, with possibly shared coefficients, then we simply draw from a multivariate normal distribution with mean the overall parameter vector, $\mathbf{b}$, and variance-covariance matrix, $\mathbf{V}$. Repeating this procedure a total of M times, we then construct 95% confidence intervals as before.

4. Time-dependent effects

The occurrence of time-dependent effects or non-proportional hazards are commonplace in medical research. For example in register based studies, where follow-up is often over many years, it is extremely important to investigate the presence of time-dependency in covariate effects. Incorporating time-dependent effects within a multi-state model framework has received very little attention. In our parametric transition specific approach, we extend the framework to allow time-dependent effects, in either the separate modelling approach, or the combined models fitted to the stacked data. The ease of incorporating time-dependent effects is a distinct advantage of the parametric approach. We therefore have for the kth transition,

$$h_k(t) = h_{0k}(t) \exp\left[X_k \beta_k(t)\right] \tag{7}$$

where $\beta_k(t)$ is now allowed to vary over time through some parametric function, relaxing the proportionality assumption. For example, for a Weibull model of the form, with baseline hazard function,

$$h_0(t) = \lambda \gamma t^{\gamma-1} \tag{8}$$

we can include covariates in the linear predictor of the shape parameter, γ, on the log scale. Similarly for the generalised gamma accelerated failure time model, we can allow the three parameters to vary depending on covariates. The most flexible model we have currently implemented within the associated Stata package is the Royston-Parmar survival model. In this model, time-dependent covariate effects can be specified by forming an interaction between the covariate and a restricted cubic spline function of log time [15]. An example can be found in the appendix. Such flexibility means that we can fit a model with any particular transition containing time-dependent effects, and then use the general simulation approach described above to calculate transition probabilities. Obtaining complex predictions is not that much harder when incorporating time-dependent effects, because we are using a parametric approach.

5. Extended predictions

When calculating transition probabilities using the general simulation algorithm described above, we generate a complete dataset of N individuals, where we know each observations exact transition history. Once the data has been generated, we can therefore use it to calculate a wide variety of clinically useful measures, with little extra effort. In this section we describe the calculations of length of stay, and describe differences and ratios of statistics, in order to illustrate covariate impact.

5.1. Length of stay

Following Grand and Putter (2015) [20], the residual restricted expected length of stay in state b during the time period from s to τ, conditional on the patient being alive at time s and in state a (non-absorbing), which we refer to as length of stay for simplicity, is defined as

$$e_{ab}(s) = \int_s^\tau P(Y(u) = b|Y(s) = a)du \tag{9}$$

which defines the amount of time spent in state b, starting in state a at time s, up until time τ. If $\tau = \infty$, and state $a = b$ is a healthy state and all possible next states are deaths, then Equation (9) represents life expectancy [20]. This can be viewed as the multi-state extension of restricted mean survival [22]. Through the simulation

approach described above, we can easily calculate mean length of stay using the observed simulated data, by simply calculating the mean length of time spent in the target state b, truncating any observation $> \tau$ at τ. Furthermore, confidence intervals are also easily calculated by repeated simulations. Arguably, the most sensible choice for τ is the maximum observed event time seen in the dataset under analysis, though a smaller value may be preferable as there is often a lack of data in the tail, and for some covariate combinations you would be extrapolating.

5.2. Differences and ratios of probabilities and length of stay

To illustrate the impact of different covariate levels, we can calculate differences and ratios between specified covariate patterns of quantities of interest, such as transition probabilities, or length of stay. Given two different covariate patterns, X_1 and X_2, we can calculate the difference between transition probabilities as,

$$P(Y(t) = b|Y(s) = a, X_1) - P(Y(t) = b|Y(s) = a, X_2) \tag{10}$$

Similarly, we can calculate difference in length of stays as,

$$e_{ab}(s, X_1) - e_{ab}(s, X_2) = \int_s^\tau P(Y(u) = b|Y(s) = a, X_1)du - \int_s^\tau P(Y(u) = b|Y(s) = a, X_2)du \tag{11}$$

Alternatively, we may be interested in the ratio between the quantities, where the ratio of transition probabilities for covariate patterns X_1 and X_2 is,

$$\frac{P(Y(t) = b|Y(s) = a, X_1)}{P(Y(t) = b|Y(s) = a, X_2)} \tag{12}$$

and equivalently for length of stay,

$$\frac{e_{ab}(s, X_1)}{e_{ab}(s, X_2)} = \frac{\int_s^\tau P(Y(u) = b|Y(s) = a, X_1)du}{\int_s^\tau P(Y(u) = b|Y(s) = a, X_2)du} \tag{13}$$

All of the quantities shown in Equations (10) to (13) are calculated simply by simulating two populations, both of size N, the first with covariate pattern X_1 and the second with covariate pattern X_2. Uncertainty is then obtained by conducting repeated simulations, drawing coefficients from the multivariate normal distribution with mean the overall parameter vector, $\mathbf{b}$, and variance-covariance matrix, $\mathbf{V}$. Repeating this procedure a total of M times, we then construct 95% confidence intervals by calculating the variance of our M estimates.

6. Modelling time to relapse and death in patients with primary breast cancer using a multi-state model

To illustrate our approach, we use data from 2,982 patients with primary breast cancer, where we have information on the time to relapse and the time to death. All patients begin in the initial post-surgery state, which is defined as the time of primary surgery, and can then move to a relapse state, or a dead state, and can also die after relapse. This is the commonly known illness-death model, shown in Figure 1.

Of the 2,982 patients starting in the healthy state, 1,518 of them went on to experience a relapse, 195 died without relapse, 1,075 died after experiencing a relapse. Covariates of interest include age at primary surgery, tumour size (three classes; $\leq$ 20mm, 20-50mm, $>$ 50mm), number of positive nodes, progesterone level (fmol/l), and whether patients were on hormonal therapy (binary, yes/no). In all analyses we use a transformation of progesterone level
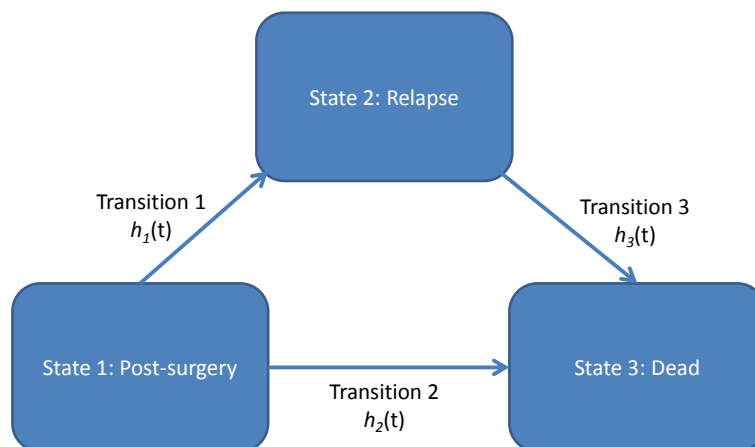
Figure 1. Illness-death model for the primary breast cancer example.

$(\log(pgr + 1))$, as has been done previously [4]. Our initial analysis focuses on how to model the baseline hazard for each of the three transitions, without any covariates. To each transition, we apply parametric models including; exponential, Weibull, Gompertz, log-logistic, log-normal, generalised gamma, and the Royston-Parmar model with 2 to 7 degrees of freedom. In Table 1, we present the AIC and BIC for each of the models fitted, and for each transition.

We find that the best fitting model for transition 1, both in terms of the AIC and BIC is the Royston-Parmar model with 3 degrees of freedom. The best fitting model for transition 2 is the Weibull based on the AIC, and the Royston-Parmar model with 2 degrees of freedom based on the BIC. The best fitting model for transition 3 is the Royston-Parmar model with 3 degrees of freedom based on the AIC, and the log-normal model based on the BIC. The fitted cumulative hazard functions for each of the transition-specific baseline models is shown in Figure 2, overlaid on the Nelson-Aalen curves.

We can now incorporate covariates into each of the selected transition-specific distributions. We include all available covariates in each of the models, and assess the proportional hazards assumption. To do this we can fit a Cox model to each transition separately and use the Schoenfeld residuals to test for non-proportionality [23], or include an interaction with a function of time. For transition 1, we found statistically significant ($p < 0.05$) evidence against the null hypothesis of proportionality, for tumour size (both levels, ref: ≤ 20mm), and progesterone level. For transition 2, evidence of non-proportionality was not found for any covariates. For transition 3, we found evidence of non-proportionality in progesterone level only. Parameter estimates for the three-transition model are in Table 2. For simplicity, we did not investigate non-linearities in covariate effects, or interactions, however both are seamlessly incorporated into the model framework. Ideally, clinical guidance should be used to drive covariate selection, but AIC/BIC can be used as informal guides.

Given our fitted multi-state model, we can now obtain predicted transition probabilities for any covariate pattern. We illustrate this in Figure 3, showing transition probabilities for a patient aged 54, with a transformed progesterone level of 3, and in each of the tumour size groups, with varying number of positive nodes $\{0, 10, 20\}$. As tumour size increases, and number of positive nodes increase, we can see the increased probability of death across follow-up.
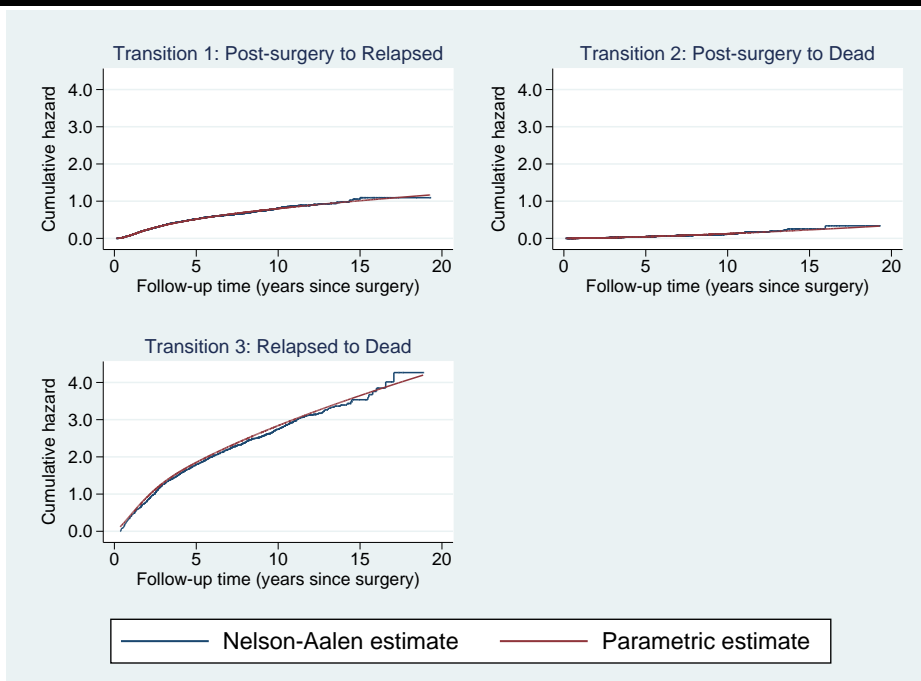
Figure 2. Best fitting parametric cumulative hazard curves overlaid on the Nelson-Aalen estimate for each transition. Transition 1 model: Royston-Parmar with 3 degrees of freedom. Transition 2 model: Weibull. Transition 3 model: Royston-Parmar with 3 degrees of freedom.

The transition probabilities shown in Figure 3 also represent state occupation probabilities, as they are calculated assuming the patient starts in state 1 at time $s = 0$. Both the stating state and entry time can be easily changed through the associated software package, to give transition probabilities for alternative settings.

Stacked transition probability plots, such as Figure 3, are an extremely useful way of showing how risk evolves over time; however, uncertainty in the estimates cannot be easily portrayed. In Figure 4, we show the predicted transition probabilities from the left panel of Figure 3, now with associated confidence intervals, using the simulation approach.

Figures 3 and 4 exemplify the ease of obtaining patient-specific, clinically useful, predictions from a fully parametric, flexible, multi-state model. We can also illustrate the impact of different covariate levels by calculating differences or ratios in transition probabilities. For example, in Figure 5 we present differences and ratios of transition probabilities for a patient aged 54, with a progesterone level of 3, and comparing the three levels of tumour size. Through simulation, we can obtain 95% confidence intervals for these contrasts.

Figure 5a illustrates the difference in transition probabilities when comparing between the smallest tumour size (≤ 20mm) and the medium sized group (>20-50mm), showing a higher probability of remaining healthy in the smaller tumour size group over time, and an increased probability of death when in the medium tumour size group. To illustrate computation time when calculating confidence intervals, to produce the data required for Figure 5a, with a default sample size of $N = 10,000$ and $M = 200$, with probabilities calculated at 100 evenly spaced times across follow-up, took 32.5 seconds on a standard HP laptop with i5 processor and 9GB of RAM. Point estimates are virtually instantaneous (under a third of a second).

We can also calculate the lengths of stay in each state, given a specific covariate pattern, such as that shown in Figure 6. Figure 6 shows that after 15 years, a patient aged 54, with progesterone level of 3 and in the smallest tumour size group, spends approximately 10 years in the healthy state, 2 years in the relapse state, and 3 years in the dead state (though as this is an absorbing state, this interpretation is less useful).

We can further show how length of stay is affected by changes in covariate patterns, for example in Figure 7 we present differences and ratios of the length of stay in each state for a patient aged 54, with a progesterone level of

3, and comparing the three levels of tumour size, with associated 95% confidence intervals for these contrasts.

We see from Figures 7a, 7c, and 7e that having a smaller tumour size indicates a greater length of stay in the healthy state, and having a larger tumour size consequently leads to a greater length of stay in the death state, over a 15 year period. Note the large confidence interval widths in Figures 7b, 7d, and 7f for states 2 and 3, due to all patients starting in the healthy state 1.

7. Discussion

Parametric survival models have a number of advantages, and we believe should be used more widely. By direct modelling of the baseline hazard function, we can gain more understanding of how complex disease processes evolve over time. Incorporating time-dependent effects, extrapolation and out of sample prediction are also much more convenient within a fully specified parametric model. Parametric survival models are of particular importance in the realm of health economic modelling [5]. There are also a variety of approaches which provide substantial flexibility in capturing complex underlying hazard functions [24, 15, 3]. In particular, the use of restricted cubic splines to capture complex hazard functions have recently been comprehensively evaluated through simulation [25, 26].

Within a multi-state framework, parametric approaches have generally only considered models which assume the same underlying distribution for all transitions. In this paper, we have introduced the use of transition-specific distributions for Markov multi-state models, allowing a wide variety of parametric distributional choices, including the combination of proportional hazards and accelerated failure time models, and spanning from the simple exponential, to the highly flexible Royston-Parmar model. We have further incorporated time-dependent effects into the modelling strategy. We then utilise a general survival simulation algorithm to calculate transition probabilities and length of stay, and associated confidence intervals, from any fitted model. This allows us to avoid the piecewise approximation of Blaser et al. (2015) [27]. We have further shown how to calculate clinically useful differences and ratios of transition probabilities and length of stays, to illustrate the impact of changing covariate levels. Hence, we can fit a complex multi-state parametric model, but still get interpretable quantities with ease. Finally, we have developed a framework allowing the flexibility of transition-specific hazard functions, but still enabled the sharing of parameters across transitions, by directly maximising the likelihood of the stacked models.

With the growing use of electronic health records, allowing the integration of clinical registries and administrative records, there are substantial opportunities to use multi-state models for expanded modelling of patient profiles across disease histories [28]. This has benefits not only for clinical effectiveness modelling, but also for cost-effectiveness modelling due to the parametric approach [29].

We illustrated our approach using a dataset of patients with primary breast cancer, who experience relapse or death, or death after relapse, i.e. an illness-death process. We chose this simple example to illustrate the framework; however, it is important to note that it generalises to any uni-directional multi-state model, such as [28, 2]. We also only discussed Markov models in this article. The extension to semi-Markov models, or the clock-reset approach, whereby we use time since entry to each state as our underlying timescale, is equally applicable, and is implemented in the associated software. We show an example in the appendix.

Our approach can be extended in a number of ways. Essentially, any functional form for the baseline hazard could be chosen, therefore we plan to incorporate the general survival modelling framework of Crowther and Lambert [3, 30]. Furthermore, an important extension is to allow reversible transitions, which will sync quite seamlessly into the modelling and simulation framework described in this article, and hence will be incorporated into the software in the near future. Incorporating frailties is another area of interest, for example, to account for recurrent events, or hierarchical structures within a dataset.

The accompanying Stata software is available from the Statistical Software Components archive, which can be

installed by typing `ssc install multistate` in Stata. This includes the `predictms` command to calculate all types of predictions described in this article, and the `stms` estimation command for fitting the transition-specific models to stacked data.

Acknowledgements

A. Example use of the `multistate` package

```
. //install the software
. ssc install multistate

.
. //load data
. use http://fmwww.bc.edu/repec/bocode/m/multistate_example,clear

.
. //data is in wide format
. list pid rf rfi os osi if pid==1 | pid==1371, sepby(pid) noobs
```

pid	rf	rfi	os	osi
1	59.1	0	59.1	alive
1371	16.6	1	24.3	deceased

```
.
. //reshape to stacked format
. msset, id(pid) states(rfi osi) times(rf os)

.
. //each patient now has a row of data for each transition they are at risk for
. list pid rf rfi os osi if pid==1 | pid==1371, sepby(pid) noobs
```

pid	rf	rfi	os	osi
1	59.1	0	59.1	alive
1	59.1	0	59.1	alive
1371	16.6	1	24.3	deceased
1371	16.6	1	24.3	deceased
1371	16.6	1	24.3	deceased

```
.
. //msset creates internal variables for use in subsequent analyses
. list pid _start _stop _from _to _status _trans if pid==1 | pid==1371
```

	pid	_start	_stop	_from	_to	_status	_trans
1.	1	0	59.104721	1	2	0	1
2.	1	0	59.104721	1	3	0	2
2765.	1371	0	16.558521	1	2	1	1
2766.	1371	0	16.558521	1	3	0	2
2767.	1371	16.558521	24.344969	2	3	1	3

```
.
. //store default transition matrix
. mat tmat = r(transmatrix)

.
. //declare as survival data (changing timescale from months to years)
. stset _stop, enter(_start) failure(_status==1) scale(12)

     failure event:  _status == 1
obs. time interval:  (0, _stop]
 enter on or after:  time _start
 exit on or before:  failure
    t for analysis:  time/12

─────────────────────────────────────────────────────────────────────────
        7482  total observations
           0  exclusions
─────────────────────────────────────────────────────────────────────────
        7482  observations remaining, representing
        2790  failures in single-record/single-failure data
   38474.539  total analysis time at risk and under observation
                                            at risk from t =          0
                                    earliest observed entry t =          0
                                       last observed exit t =   19.28268

.
. //create indicator variables for tumour size
. tab size, gen(sz)
```

Tumour size, 3 classes (t)	Freq.	Percent	Cum.
<=20 mm	3,339	44.63	44.63
>20-50mmm	3,327	44.47	89.09
>50 mm	816	10.91	100.00
Total	7,482	100.00	

```
.
. //fit final models to each of the three transitions and store the model object
> ts
. //trans1
. stpm2 age sz2 sz3 nodes pr_1 hormon if _trans==1, df(3) scale(h)        ///
>         tvc(sz2 sz3 pr_1) dftvc(1)

Iteration 0:   log likelihood = -3522.9535
Iteration 1:   log likelihood = -3479.1473
Iteration 2:   log likelihood = -3476.6534
Iteration 3:   log likelihood = -3476.6455
Iteration 4:   log likelihood = -3476.6455

Log likelihood = -3476.6455                     Number of obs      =      2,982

─────────────────────────────────────────────────────────────────────────
          |     Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
```

```
  xb
         age |  -.0062709   .0021004    -2.99   0.003    -.0103875   -.0021543
         sz2 |   .4777289   .0634816     7.53   0.000     .3533073    .6021505
         sz3 |    .744544   .0904352     8.23   0.000     .5672943    .9217937
       nodes |   .0784025   .0045454    17.25   0.000     .0694937    .0873113
        pr_1 |  -.0783066   .0122404    -6.40   0.000    -.1022973   -.0543159
      hormon |  -.0797426   .0824504    -0.97   0.333    -.2413424    .0818572
       _rcs1 |   .9703563   .0472652    20.53   0.000     .8777182    1.062994
       _rcs2 |   .3104222   .0218912    14.18   0.000     .2675162    .3533282
       _rcs3 |  -.0176099   .0114839    -1.53   0.125    -.0401179    .0048982
   _rcs_sz21 |  -.1740546   .0446893    -3.89   0.000     -.261644   -.0864652
   _rcs_sz31 |  -.2669255   .0616161    -4.33   0.000    -.3876909   -.1461601
   _rcs_pr_11 |    .072824   .0086399     8.43   0.000     .0558901    .0897578
       _cons |  -.9480559   .1266088    -7.49   0.000    -1.196205   -.6999071
```

. est store m1

.

. //trans 2

. streg age sz2 sz3 nodes pr_1 hormon if _trans==2, dist(weib) nohr

```
        failure _d:  _status == 1
  analysis time _t:  _stop/12
  enter on or after:  time _start
```

Fitting constant-only model:

```
Iteration 0:   log likelihood = -789.49708
Iteration 1:   log likelihood = -772.22637
Iteration 2:   log likelihood = -771.90383
Iteration 3:   log likelihood = -771.90373
Iteration 4:   log likelihood = -771.90373
```

Fitting full model:

```
Iteration 0:   log likelihood = -771.90373
Iteration 1:   log likelihood = -759.32799
Iteration 2:   log likelihood = -585.02458
Iteration 3:   log likelihood = -580.52674
Iteration 4:   log likelihood = -580.44872
Iteration 5:   log likelihood =  -580.4487
```

Weibull regression -- log relative-hazard form

```
No. of subjects =        2,982                   Number of obs    =        2,982
No. of failures =          195
Time at risk    =  17203.80014
                                                 LR chi2(6)       =       382.91
Log likelihood  =    -580.4487                   Prob > chi2      =       0.0000
```

```
          _t |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
         age |   .1250736   .0079699    15.69   0.000     .1094529    .1406942
         sz2 |   .1615513   .1614483     1.00   0.317    -.1548816    .4779841
         sz3 |   .4153083   .2332724     1.78   0.075    -.0418973    .8725138
       nodes |   .0439416   .0182545     2.41   0.016     .0081633    .0797198
        pr_1 |   .0223507   .0334238     0.67   0.504    -.0431587    .0878601
      hormon |  -.1399098   .2291893    -0.61   0.542    -.5891126     .309293
       _cons |  -14.02238   .6218293   -22.55   0.000    -15.24114   -12.80362

       /ln_p |   .5106518   .0572511     8.92   0.000     .3984416     .622862

           p |   1.666377    .095402                      1.489502    1.864256
```

```
            1/p  |    .6001043   .0343567                     .5364071    .6713655
            ─────────────────────────────────────────────────────────────────────

. est store m2

.
. //trans 3
. stpm2 age sz2 sz3 nodes pr_1 hormon if _trans==3, df(3) scale(h)          ///
>         tvc(pr_1) dftvc(1)
note: delayed entry models are being fitted

Iteration 0:   log likelihood = -943.78807
Iteration 1:   log likelihood = -930.00714
Iteration 2:   log likelihood =  -929.1202
Iteration 3:   log likelihood = -929.11658
Iteration 4:   log likelihood = -929.11658

Log likelihood = -929.11658                    Number of obs     =      1,518
```

	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
xb						
age	.0049441	.0024217	2.04	0.041	.0001977	.0096906
sz2	.1653563	.0712326	2.32	0.020	.025743	.3049696
sz3	.3243048	.0992351	3.27	0.001	.1298075	.5188021
nodes	.0297031	.0057735	5.14	0.000	.0183873	.0410189
pr_1	-.1843876	.0211383	-8.72	0.000	-.225818	-.1429572
hormon	.0315634	.0976384	0.32	0.746	-.1598045	.2229312
_rcs1	.5057489	.0581187	8.70	0.000	.3918383	.6196595
_rcs2	.1035699	.03143	3.30	0.001	.0419681	.1651716
_rcs3	-.0100584	.0117741	-0.85	0.393	-.0331352	.0130185
_rcs_pr11	.0636225	.0121503	5.24	0.000	.0398085	.0874366
_cons	.391217	.1659763	2.36	0.018	.0659094	.7165246

```
. est store m3

.
. //calculate transition probabilities, starting in state 1 at time 0 for a pat
> ient age 54 (all other covariates set to 0)
. predictms, transmatrix(tmat) models(m1 m2 m3) at(age 50)

.
. //produce stacked graph
. predictms, transmatrix(tmat) models(m1 m2 m3) at(age 50) graph

.
. //calculate confidence intervals
. predictms, transmatrix(tmat) models(m1 m2 m3) at(age 50) obs(100) n(10000) m(
> 200) ci

.
. //calculate length of stay in each state
. predictms, transmatrix(tmat) models(m1 m2 m3) at(age 50) los

.
. //calculate the difference in transition probabilities for a patient with age
>  50 compared to a patient aged 60:
. predictms, transmatrix(tmat) models(m1 m2 m3) at(age 50) at2(age 60)

.
. //calculate the ratio of transition probabilities for a patient with age 50 c
> ompared to a patient aged 60:
. predictms, transmatrix(tmat) models(m1 m2 m3) at(age 50) at2(age 60) ratio

.
```

```
. //calculate the ratio of length of stay for a patient with age 50 compared to
>  a patient aged 60, with confidence intervals:
. predictms, transmatrix(tmat) models(m1 m2 m3) at(age 50) at2(age 60) los rati
> o ci

.
. //fit a transition-specific multi-sate model, but sharing the effect of age a
> cross transitions and 3
. stms    (age sz2 sz3 nodes pr_1 hormon, model(fpm) df(3) scale(h))  ///
>         (age sz2 sz3 nodes pr_1 hormon, model(weib))          ///
>         (age sz2 sz3 nodes pr_1 hormon, model(fpm) df(3) scale(h))  ///
>         , transvar(_trans) constraints(age 1 3)
```

To fit a semi-Markov model, we simply use a different timescale, i.e. the time since entry into each state. Below we show how to do this with the `multistate` package. Essentially we use a new time variable, and specify the `reset` option of `predictms`. All other commands are as above.

```
. //install the software
. ssc install multistate
checking multistate consistency and verifying not already installed...
all files already exist and are up to date.

.
. //load data
. use http://fmwww.bc.edu/repec/bocode/m/multistate_example,clear

.
. //reshape to stacked format
. msset, id(pid) states(rfi osi) times(rf os)

.
. //store default transition matrix
. mat tmat = r(transmatrix)

.
. //generate time since entry
. gen _time = _stop - _start

.
. //declare as survival data (changing timescale from months to years)
. stset _time, failure(_status==1) scale(12)

     failure event:  _status == 1
obs. time interval:  (0, _time]
 exit on or before:  failure
    t for analysis:  time/12

_____

        7482  total observations
           0  exclusions
_____

        7482  observations remaining, representing
        2790  failures in single-record/single-failure data
   38474.539  total analysis time at risk and under observation
                                            at risk from t =          0
                                  earliest observed entry t =          0
                                     last observed exit t =   19.28268

.
. //create indicator variables for tumour size
. tab size, gen(sz)

     Tumour |
    size, 3 |
 classes (t) |      Freq.      Percent        Cum.
```

```
         ──────────────┬─────────────────────────────────
          <=20 mm       │    3,339      44.63      44.63
          >20-50mmm     │    3,327      44.47      89.09
          >50 mm        │      816      10.91     100.00
         ──────────────┼─────────────────────────────────
          Total         │    7,482     100.00
```

.
. //fit final models to each of the three transitions and store the model objects
. //trans1
. stpm2 age sz2 sz3 nodes pr_1 hormon if _trans==1, df(3) scale(h)   ///
>   tvc(sz2 sz3 pr_1) dftvc(1)

```
Iteration 0:   log likelihood = -3522.9535
Iteration 1:   log likelihood = -3479.1473
Iteration 2:   log likelihood = -3476.6534
Iteration 3:   log likelihood = -3476.6455
Iteration 4:   log likelihood = -3476.6455

Log likelihood = -3476.6455                 Number of obs     =      2,982

           ┌───────────────────────────────────────────────────────────────────
           │     Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
───────────┼───────────────────────────────────────────────────────────────────
xb         │
       age │  -.0062709   .0021004    -2.99   0.003    -.0103875   -.0021543
       sz2 │   .4777289   .0634816     7.53   0.000     .3533073    .6021505
       sz3 │    .744544   .0904352     8.23   0.000     .5672943    .9217937
     nodes │   .0784025   .0045454    17.25   0.000     .0694937    .0873113
      pr_1 │  -.0783066   .0122404    -6.40   0.000    -.1022973   -.0543159
    hormon │  -.0797426   .0824504    -0.97   0.333    -.2413424    .0818572
     _rcs1 │   .9703563   .0472652    20.53   0.000     .8777182    1.062994
     _rcs2 │   .3104222   .0218912    14.18   0.000     .2675162    .3533282
     _rcs3 │  -.0176099   .0114839    -1.53   0.125    -.0401179    .0048982
 _rcs_sz21 │  -.1740546   .0446893    -3.89   0.000     -.261644   -.0864652
 _rcs_sz31 │  -.2669255   .0616161    -4.33   0.000    -.3876909   -.1461601
_rcs_pr_11 │    .072824   .0086399     8.43   0.000     .0558901    .0897578
     _cons │  -.9480559   .1266088    -7.49   0.000    -1.196205   -.6999071
           └───────────────────────────────────────────────────────────────────
```

. est store m1

.
. //trans 2
. streg age sz2 sz3 nodes pr_1 hormon if _trans==2, dist(weib) nohr

```
        failure _d:  _status == 1
   analysis time _t:  _time/12

Fitting constant-only model:

Iteration 0:   log likelihood = -789.49708
Iteration 1:   log likelihood = -772.22637
Iteration 2:   log likelihood = -771.90383
Iteration 3:   log likelihood = -771.90373
Iteration 4:   log likelihood = -771.90373

Fitting full model:

Iteration 0:   log likelihood = -771.90373
Iteration 1:   log likelihood = -759.32799
Iteration 2:   log likelihood = -585.02458
Iteration 3:   log likelihood = -580.52674
Iteration 4:   log likelihood = -580.44872
Iteration 5:   log likelihood =  -580.4487

Weibull regression -- log relative-hazard form
```

```
No. of subjects =        2,982          Number of obs     =        2,982
No. of failures =          195
Time at risk     =   17203.80014
                                        LR chi2(6)        =       382.91
Log likelihood   =    -580.4487         Prob > chi2       =       0.0000
```

_t	Coef.	Std. Err.	z	P>|z|	[95% Conf. Interval]	
age	.1250736	.0079699	15.69	0.000	.1094529	.1406942
sz2	.1615513	.1614483	1.00	0.317	-.1548816	.4779841
sz3	.4153083	.2332724	1.78	0.075	-.0418973	.8725138
nodes	.0439416	.0182545	2.41	0.016	.0081633	.0797198
pr_1	.0223507	.0334238	0.67	0.504	-.0431587	.0878601
hormon	-.1399098	.2291893	-0.61	0.542	-.5891126	.309293
_cons	-14.02238	.6218293	-22.55	0.000	-15.24114	-12.80362
/ln_p	.5106518	.0572511	8.92	0.000	.3984416	.622862
p	1.666377	.095402			1.489502	1.864256
1/p	.6001043	.0343567			.5364071	.6713655

```
. est store m2

.
. //trans 3
. stpm2 age sz2 sz3 nodes pr_1 hormon if _trans==3, df(3) scale(h)        ///
>         tvc(pr_1) dftvc(1)

Iteration 0:   log likelihood = -2083.3928
Iteration 1:   log likelihood =  -2075.987
Iteration 2:   log likelihood = -2075.9448
Iteration 3:   log likelihood = -2075.9448

Log likelihood = -2075.9448                    Number of obs     =        1,518
```

	Coef.	Std. Err.	z	P>|z|	[95% Conf. Interval]	
xb						
age	.0052585	.0024332	2.16	0.031	.0004895	.0100276
sz2	.212097	.0709341	2.99	0.003	.0730688	.3511253
sz3	.4260375	.0980746	4.34	0.000	.2338147	.6182602
nodes	.0347199	.0055692	6.23	0.000	.0238045	.0456354
pr_1	-.1644984	.0172515	-9.54	0.000	-.1983108	-.130686
hormon	.1006494	.0966689	1.04	0.298	-.0888182	.290117
_rcs1	1.228126	.0606233	20.26	0.000	1.109307	1.346946
_rcs2	-.006099	.0409719	-0.15	0.882	-.0864024	.0742045
_rcs3	.0629436	.0179447	3.51	0.000	.0277726	.0981146
_rcs_pr_11	.063054	.0157993	3.99	0.000	.0320879	.0940202
_cons	-1.01157	.1495324	-6.76	0.000	-1.304648	-.7184917

```
. est store m3

.
. //calculate transition probabilities, starting in state 1 at time 0 for a pat
> ient age 54 (all other covariates set to 0)
. predictms, transmatrix(tmat) models(m1 m2 m3) at(age 50) reset
```

References

1. Putter H, Fiocco M, Geskus RB. Tutorial in biostatistics: competing risks and multi-state models. *Stat Med* 2007; **26**(11):2389–2430, doi:10.1002/sim.2712. URL `http://dx.doi.org/10.1002/sim.2712`.

2. de Wreede LC, Fiocco M, Putter H. mstate: An R Package for the Analysis of Competing Risks and Multi-State Models. *Journal of Statistical Software* 1 2011; **38**(7):1–30. URL `http://www.jstatsoft.org/v38/i07`.

3. Crowther MJ, Lambert PC. A general framework for parametric survival analysis. *Stat Med* Dec 2014; **33**(30):5280–5297, doi:10.1002/sim.6300. URL `http://dx.doi.org/10.1002/sim.6300`.

4. Royston P, Lambert PC. *Flexible Parametric Survival Analysis Using Stata: Beyond the Cox Model*. Stata Press, 2011.

5. 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):743–754, doi:10.1177/0272989X12472398. URL `http://dx.doi.org/10.1177/0272989X12472398`.

6. Andersson TML, Dickman PW, Eloranta S, Lambe M, Lambert PC. Estimating the loss in expectation of life due to cancer using flexible parametric survival models. *Stat Med* Aug 2013; doi:10.1002/sim.5943. URL `http://dx.doi.org/10.1002/sim.5943`.

7. Jackson C. flexsurv: A platform for parametric survival modeling in r. *Journal of Statistical Software* 2016; **70**(1):1–33, doi:10.18637/jss.v070.i08. URL `https://www.jstatsoft.org/index.php/jss/article/view/v070i08`.

8. Omar RZ, Stallard N, Whitehead J. A parametric multistate model for the analysis of carcinogenicity experiments. *Lifetime Data Anal* 1995; **1**(4):327–346.

9. Jackson CH. Multi-state models for panel data: the msm package for R. *Journal of Statistical Software* 2011; **38**(8):1–29.

10. Titman AC. Flexible nonhomogeneous Markov models for panel observed data. *Biometrics* Sep 2011; **67**(3):780–787, doi:10.1111/j.1541-0420.2010.01550.x. URL `http://dx.doi.org/10.1111/j.1541-0420.2010.01550.x`.

11. Król A, Saint-Pierre P. SemiMarkov: An R Package for Parametric Estimation in Multi-State Semi-Markov Models. *Journal of Statistical Software* 2015; **66**(1):1–16.

12. Hsieh HJ, Chen THH, Chang SH. Assessing chronic disease progression using non-homogeneous exponential regression Markov models: an illustration using a selective breast cancer screening in Taiwan. *Statistics in medicine* 2002; **21**(22):3369–3382.

13. Hinchliffe SR, Scott DA, Lambert PC. Flexible parametric illness-death models. *Stata Journal* 2013; **13**(4):759–775.

14. Cox C, Chu H, Schneider MF, Munoz A. Parametric Survival Analysis and Taxonomy of Hazard Functions for the Generalized Gamma Distribution. *Stat Med* 2007; **26**(23):4352–4374, doi:10.1002/sim.2836.

15. Royston P, Parmar MKB. Flexible Parametric Proportional Hazards and Proportional Odds Models for Censored Survival Data, with Application to Prognostic Modelling and Estimation of Treatment Effects. *Stat Med* 2002; **21**(15):2175–2197.

16. Fiocco M, Putter H, van Houwelingen HC. Reduced-rank proportional hazards regression and simulation-based prediction for multi-state models. *Stat Med* Sep 2008; **27**(21):4340–4358, doi:10.1002/sim.3305. URL `http://dx.doi.org/10.1002/sim.3305`.

17. Touraine C, Helmer C, Joly P. Predictions in an illness-death model. *Statistical methods in medical research* 2013; .

18. Iacobelli S, Carstensen B. Multiple time scales in multi-state models. *Stat Med* Dec 2013; **32**(30):5315–5327, doi:10.1002/sim.5976. URL `http://dx.doi.org/10.1002/sim.5976`.

19. Crowther MJ, Lambert PC. Simulating biologically plausible complex survival data. *Stat Med* 2013; **32**(23):4118–4134, doi:10.1002/sim.5823. URL `http://dx.doi.org/10.1002/sim.5823`.

20. Grand MK, Putter H. Regression models for expected length of stay. *Stat Med* Oct 2015; doi:10.1002/sim.6771. URL `http://dx.doi.org/10.1002/sim.6771`.

21. Hinchliffe SR, Crowther MJ, Phillips RS, Sutton AJ. Using meta-analysis to inform the design of subsequent studies of diagnostic test accuracy. *Res Synth Methods* Jun 2013; **4**(2):156–168, doi:10.1002/jrsm.1066. URL `http://dx.doi.org/10.1002/jrsm.1066`.

22. Royston P, Parmar MKB. The use of restricted mean survival time to estimate the treatment effect in randomized clinical trials when the proportional hazards assumption is in doubt. *Stat Med* Aug 2011; **30**(19):2409–2421, doi:10.1002/sim.4274. URL `http://dx.doi.org/10.1002/sim.4274`.

23. Schoenfeld D. Partial Residuals for the Proportional Hazards Regression Model. *Biometrika* 1982; **69**(1):239–241.

24. Kooperberg C, Clarkson DB. Hazard regression with interval-censored data. *Biometrics* 1997; **53**(4):1485–1494.

25. Rutherford MJ, Crowther MJ, Lambert PC. The use of restricted cubic splines to approximate complex hazard functions in the analysis of time-to-event data: a simulation study. *J Statist Comput Simulation* 2015; **85**(4):777–793, doi:10.1080/00949655.2013.845890. URL `http://www.tandfonline.com/doi/abs/10.1080/00949655.2013.845890`.

26. Bower H, Crowther MJ, Rutherford MJ, Andersson TML, Clements M, Liu X, Dickman PW, Lambert PC. Capturing simple and complex time-dependent effects using flexible parametric survival models: a simulation study. *Under revision* ; .

27. Blaser N, Vizcaya LS, Estill J, Zahnd C, Kalesan B, Egger M, Gsponer T, Keiser O. gems: an r package for simulating from disease progression models. *Journal of statistical software* 2015; **64**(10):1.

28. 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* 2015; :0962280215578 777.

29. Asaria M, Walker S, Palmer S, Gale CP, Shah AD, Abrams KR, Crowther M, Manca A, Timmis A, Hemingway H, *et al.*. Using electronic health records to predict costs and outcomes in stable coronary artery disease. *Heart* 2016; :heartjnl–2015.

30. Crowther MJ, Lambert PC. stgenreg: A Stata package for the general parametric analysis of survival data. *J Stat Softw* 2013; **53**(12).

Table 1. AIC and BIC for each parametric model fitted to each transition separately.

Transition	Criteria	Exp.	Weibull	Gompertz	log-logistic	log-normal	Gen. gamma	R-P DF=2	R-P DF=3	R-P DF=4	R-P DF=5	R-P DF=6	R-P DF=7
1	AIC	7781.0	7766.1	7672.9	7654.8	7565.5	7503.4	7509.8	7499.2	7500.4	7500.6	7500.4	7501.7
	BIC	7786.9	7778.0	7684.8	7666.7	7577.4	7521.2	7525.8	7520.5	7527.0	7532.6	7537.7	7544.3
2	AIC	1581.0	1547.8	1548.2	1549.1	1564.0	1548.7	1548.4	1550.2	1549.5	1548.4	1549.9	1552.4
	BIC	1586.9	1559.7	1560.0	1561.0	1575.8	1566.5	1558.2	1563.3	1565.9	1568.0	1572.8	1578.6
3	AIC	2132.3	2008.0	2022.9	2013.2	2004.5	2006.5	2005.5	2003.1	2004.3	2004.7	2005.2	2008.2
	BIC	2138.3	2019.8	2034.8	2025.0	2016.4	2024.3	2020.5	2023.0	2029.2	2034.5	2040.0	2048.1

Table 2. Model estimates from the transition-specific models fitted to the primary breast cancer dataset. Transitions 1 and 3 were Royston-Parmar models, transition 2 is a Weibull.

Covariate	Transition 1: Post-surgery to relapse			Transition 2: Post-surgery to dead			Transition 3: Relapse to dead		
	Estimate	Std. Err.	95% CI	Estimate	Std. Err.	95% CI	Estimate	Std. Err.	95% CI
Age (years)	-0.006	0.002	-0.010 -0.002	0.125	0.008	0.109 0.141	0.005	0.002	0.000 0.010
Tumour size >20-50mm	0.478	0.063	0.353 0.602	0.162	0.161	-0.155 0.478	0.170	0.071	0.030 0.309
Tumour size >50mm	0.745	0.090	0.567 0.922	0.415	0.233	-0.042 0.873	0.325	0.099	0.130 0.520
Number of nodes	0.078	0.005	0.069 0.087	0.044	0.018	0.008 0.080	0.030	0.006	0.018 0.041
Progesterone level	-0.078	0.012	-0.102 -0.054	0.022	0.033	-0.043 0.088	-0.183	0.021	-0.224 -0.141
Hormone therapy	-0.080	0.082	-0.241 0.082	-0.140	0.229	-0.589 0.309	0.027	0.098	-0.165 0.219
Spline 1	0.970	0.047	0.878 1.063	†0.511	0.057	0.398 0.623	0.512	0.058	0.398 0.625
Spline 2	0.310	0.022	0.268 0.353				0.106	0.032	0.044 0.168
Spline 3	-0.018	0.011	-0.040 0.005				-0.010	0.012	-0.033 0.013
Spline 1a * (Tumour size >20-50mm)	-0.174	0.045	-0.262 -0.086						
Spline 1b * (Tumour size >50mm)	-0.267	0.062	-0.388 -0.146						
Spline 1c * progesterone level	0.073	0.009	0.056 0.090				0.063	0.012	0.039 0.087
Intercept	-0.948	0.127	-1.196 -0.700	-14.022	0.622	-15.241 -12.804	0.377	0.165	0.052 0.701

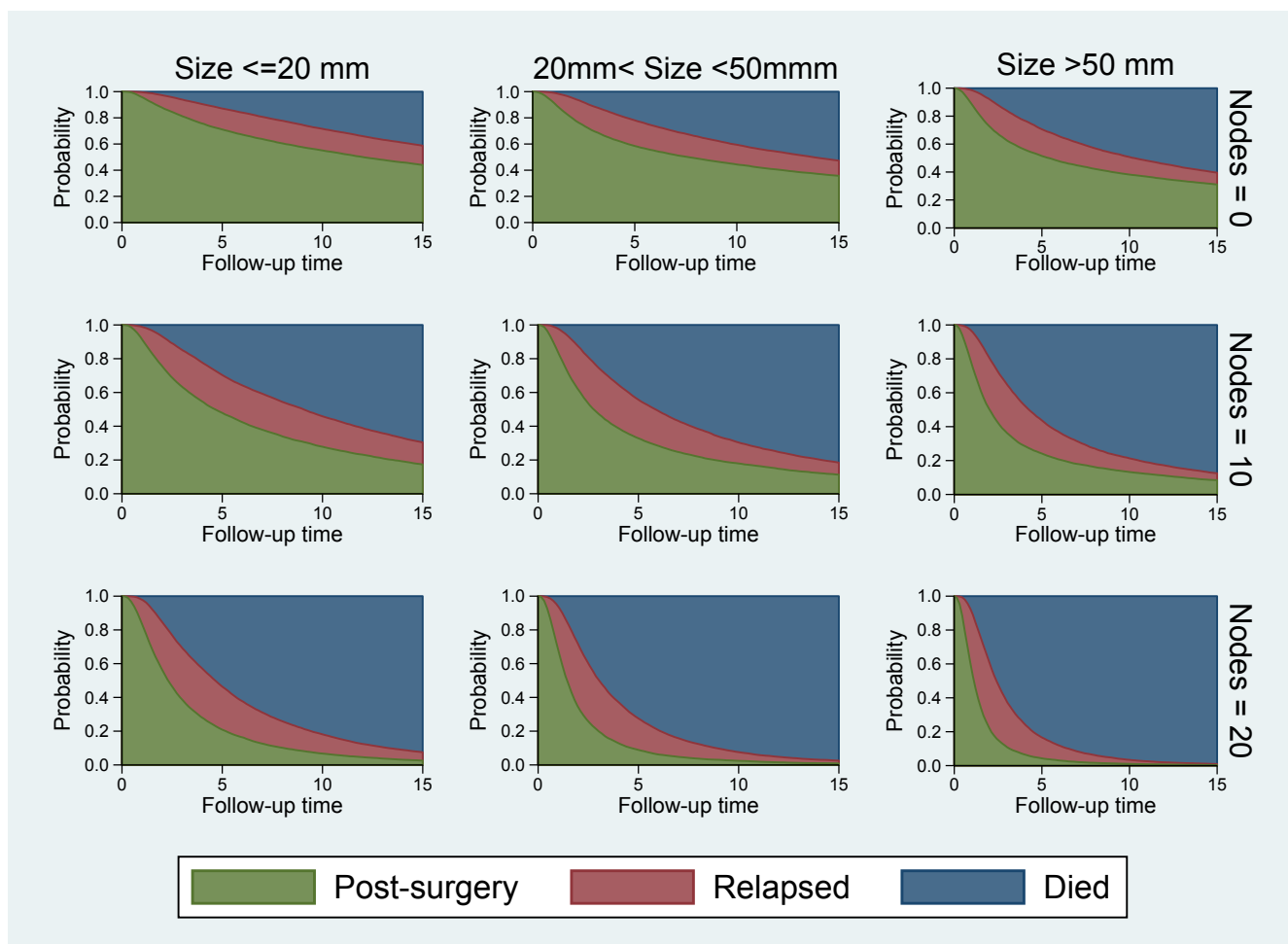†$\log(\gamma)$: the log shape parameter from the Weibull model for transition 2.

Figure 3. Probability of being in each state for a patient aged 54, with progesterone level (transformed scale) of 3, varying across tumour size and number of positive nodes.
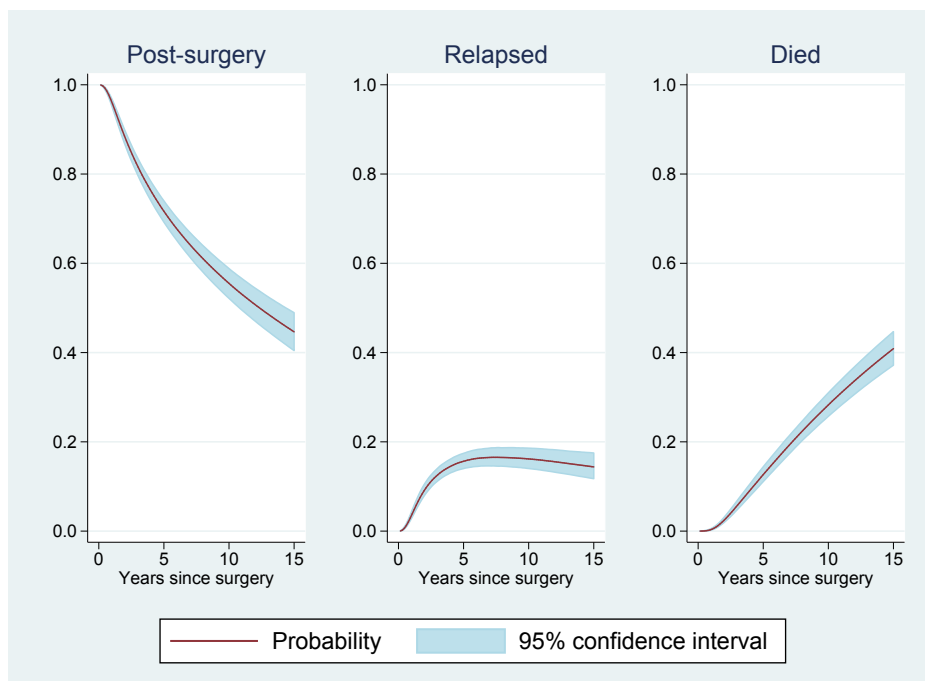
Figure 4. Probability of being in each state for a patient aged 54, size ≤20 mm, with progesterone level (transformed scale) of 3, and associated confidence intervals.
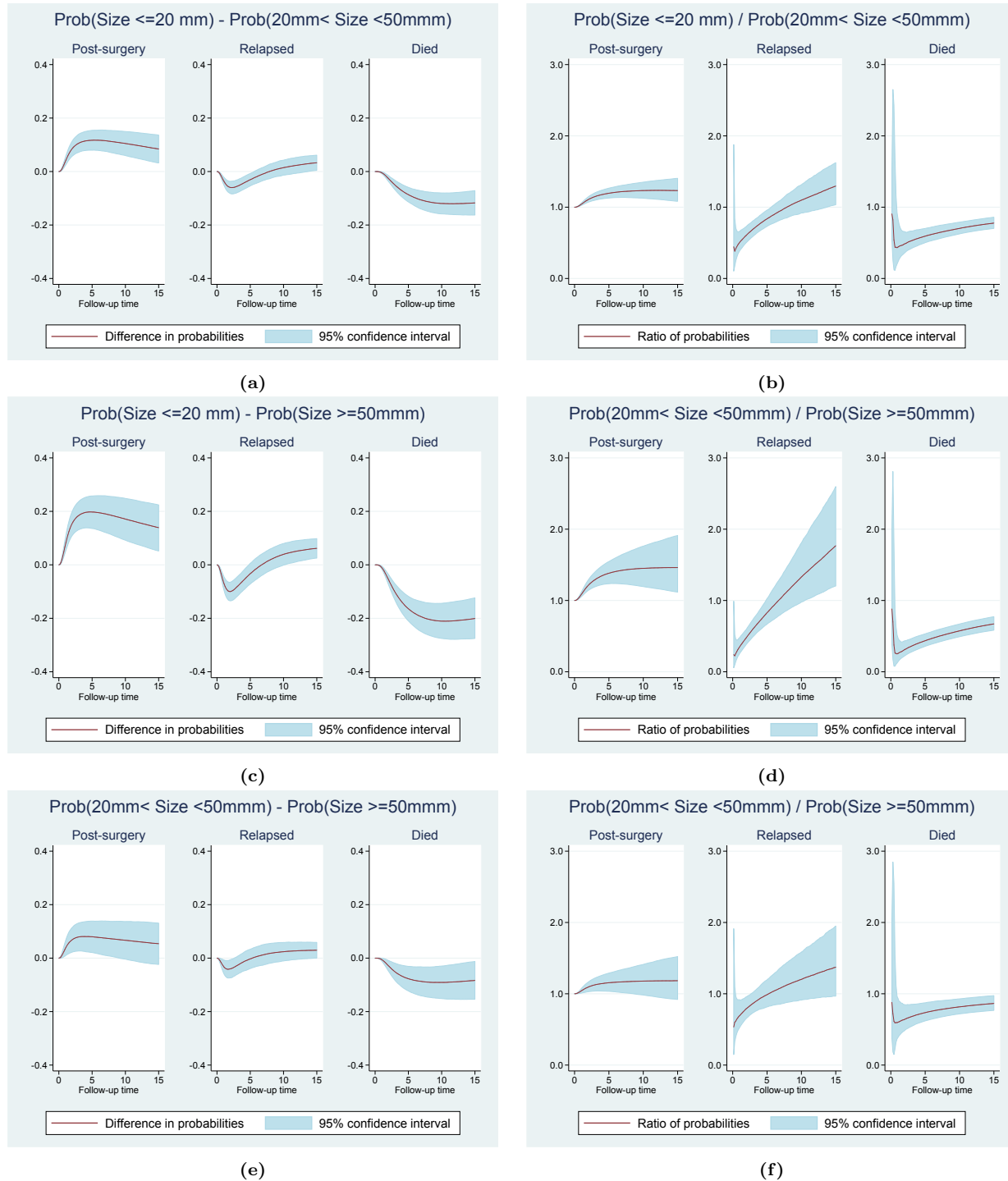
Figure 5. Differences (5a, 5c, 5e) and ratios (5b, 5d, 5f) of probabilities of being in each state for a patient aged 54, progesterone level (transformed scale) of 3, and comparing the three levels of tumour size. Associated 95% confidence intervals are also shown.
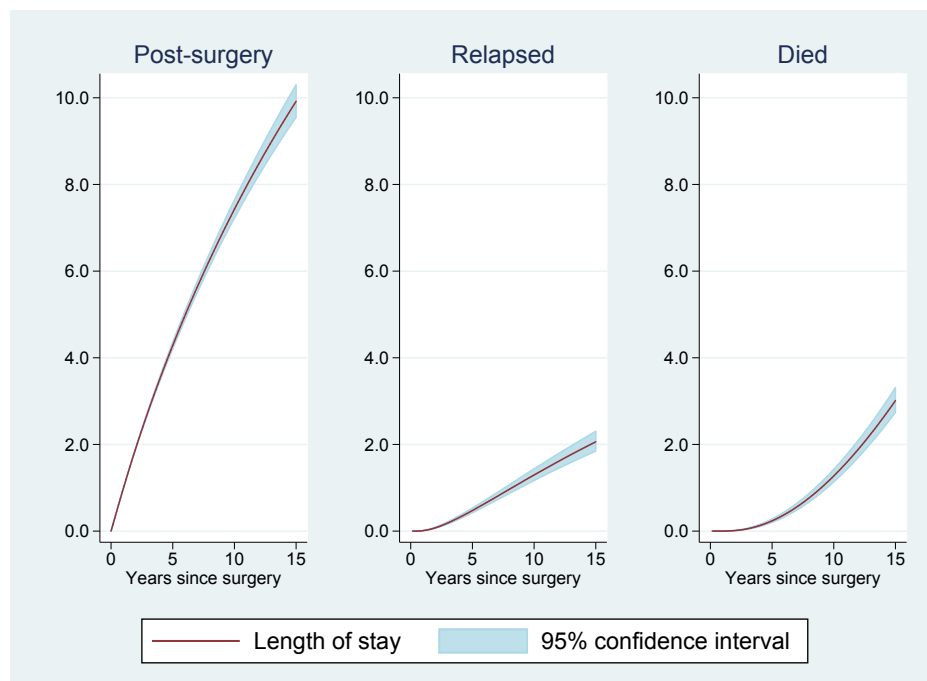
Figure 6. Length of stay in each state for a patient aged 54, size ≤20 mm, with progesterone level (transformed scale) of 3, and the smallest tumour size, with associated confidence intervals.
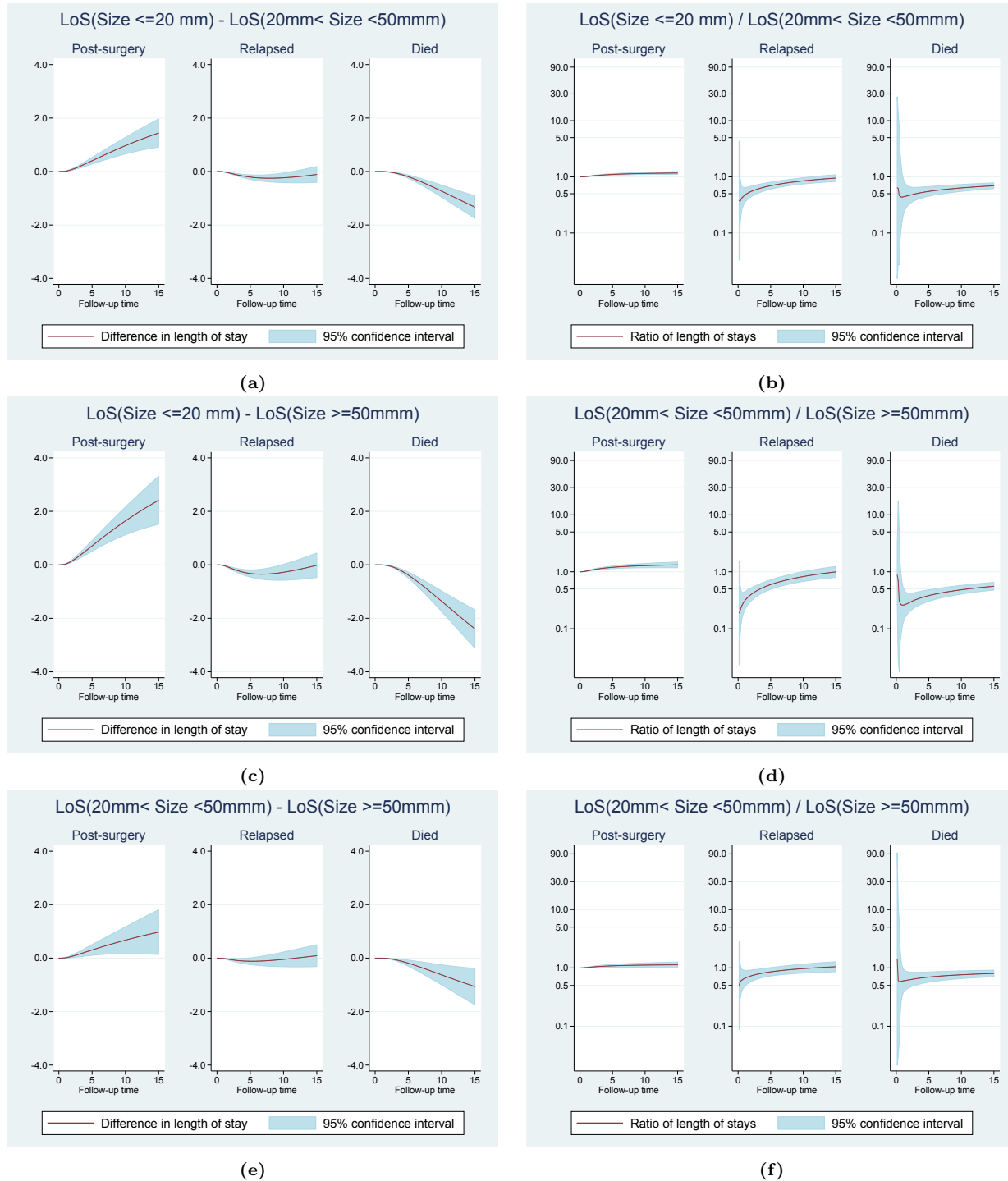
Figure 7. Differences (7a, 7c, 7e) and ratios (7b, 7d, 7f) of length of stay in each state for a patient aged 54, progesterone level (transformed scale) of 3, and comparing the three levels of tumour size. Associated 95% confidence intervals are also shown.