Title

Extrapolation of survival curves using standard parametric models and flexible parametric spline models: comparisons in large registry cohorts with advanced cancer

Running head

Extrapolation of survival curves

Authors

Jodi Gray, MPH [1]
Phone: +61 8 7421 9818
Email: jodi.gray@flinders.edu.au

Thomas Sullivan, PhD [2,3]
Phone: +61 8 8128 4419
Email: thomas.sullivan@sahmri.com

Nicholas R. Latimer, PhD [4]
Phone: +44 0 114 222 0821
Email: n.latimer@sheffield.ac.uk

Amy Salter, PhD [3]
Phone: +61 8 8313 4619
Email: amy.salter@adelaide.edu.au

Michael J. Sorich, PhD [1]
Phone: +61 8 8204 6682
Email: michael.sorich@flinders.edu.au

Robyn L. Ward, PhD [5]
Phone: +61 2 8627 5492
Email: robyn.ward@sydney.edu.au

Jonathan Karnon, PhD [1]
Phone: +61 8 7421 9821
Email: jonathan.karnon@flinders.edu.au

Affiliations

[1] Flinders Health and Medical Research Institute (FHMRI), Flinders University, Australia
[2] South Australian Health and Medical Research Institute (SAHMRI), Australia
[3] School of Public Health, The University of Adelaide, Australia
[4] School of Health and Related Research (ScHARR), University of Sheffield, UK
[5] Faculty of Medicine and Health, University of Sydney, Australia

Mailing addresses

[1] Flinders University, GPO Box 2100, Adelaide, South Australia 5001 Australia
[2] SAHMRI, PO Box 11060, Adelaide, South Australia 5001 Australia
[3] School of Public Health, AHMS Building, The University of Adelaide, South Australia 5005 Australia

[4] Health Economics and Decision Science (HEDS), School of Health and Related Research (ScHARR), University of Sheffield, Regent Court, 30 Regent Street, Sheffield, S1 4DA, United Kingdom
[5] A14 Quadrangle, The University of Sydney, New South Wales 2006 Australia

Corresponding author

Jodi Gray
Mailing address: Flinders University, GPO Box 2100, Adelaide, South Australia 5001 Australia
Phone: +61 8 7421 9818
Email: jodi.gray@flinders.edu.au

Institution where the work was done

College of Medicine and Public Health, Flinders University, Australia

Presentations based on this work have been given at

None.

Financial support

Financial support for this study was provided entirely by a grant from the National Health and Medical Research Council (NHMRC). Additionally, Dr Nicholas Latimer was supported by the National Institute for Health Research (NIHR Post Doctoral Fellowship, Dr Nicholas Latimer, PDF-2015-08-022) and is now supported by Yorkshire Cancer Research (Award reference number S406NL). The funding agreements ensured the authors' independence in designing the study, interpreting the data, writing, and publishing the report. The views expressed in this publication are those of the author(s) and not necessarily those of the NHS, the National Institute for Health Research, the Department of Health and Social Care, or Yorkshire Cancer Research. Professor Robyn Ward chairs the Medical Services Advisory Committee for the Commonwealth of Australia and the views in this paper do not reflect those of the Commonwealth.

Dr Michael Sorich reports grants from Pfizer outside the submitted work. Dr Nicholas Latimer reports consultancy fees from Astra Zeneca, Bluebird Bio, Merck EMD Serono, Pierre Fabre outside the submitted work. The other authors declare that there are no conflicts of interest.

Word count

6013

Abstract

Background. It is often important to extrapolate survival estimates beyond the limited follow-up times of clinical trials. Extrapolated survival estimates can be highly sensitive to model choice, thus appropriate model selection is crucial. Flexible parametric spline models have been suggested as an alternative to standard parametric models, however their ability to extrapolate is not well understood. **Aim.** To determine how well standard parametric and flexible parametric spline models predict survival when fitted to registry cohorts with artificially right-censored follow-up times. **Methods.** Adults with advanced breast, colorectal, small cell lung, non-small cell lung, or pancreatic cancer with a potential follow-up time of 10 years were selected from the SEER 1973-2015 registry dataset. Patients were classified into 15 cohorts by cancer and age group at diagnosis (18-59, 60-69, 70+ years). Follow-up times for each cohort were right-censored at 20%, 35% and 50% survival. Standard parametric models (exponential, Weibull, Gompertz, log-logistic, log-normal, generalized gamma) and spline models (proportional hazards, proportional odds, normal/probit) were fitted to the 10-year dataset and the three right-censored datasets. Predicted 10-year restricted mean survival time and percentage surviving at 10 years were compared to the observed values. **Results.** Across all datasets the spline odds and spline normal models most frequently gave accurate predictions of 10-year survival outcomes. Visually, spline models tended to demonstrate better fit to the observed hazard functions than standard parametric models, both in the censored and 10-year data. **Conclusions.** In these cohorts, where there was little uncertainty in the observed data, the spline models performed well when extrapolating beyond the observed data. Spline models should be routinely included in the set of models that are fitted when extrapolating cancer survival data.

Key words

oncology, extrapolation, model selection, survival analysis, cost-effectiveness analysis, modelling, restricted mean survival time, overall survival, prediction, censoring, parametric models, flexible parametric spline models, Royston and Parmar spline models

Introduction

Clinical trials for cancer pharmaceuticals, even those for metastatic cancer, are often analyzed and completed before all participating patients are deceased. Hence survival times for some patients are incomplete (right censored), mean (unrestricted) overall survival times cannot be calculated from observed data and extrapolation is required to estimate important differences in costs and outcomes between the intervention and comparator groups.[1,2] Extrapolated survival estimates can be highly sensitive to the choice of model.[3-6] Selecting an inappropriate model can strongly bias survival estimates and lead to inaccurate cost-effectiveness results. Thus providing adequate justification for the selected model increases the confidence of the decision maker in the validity of the analysis and hence, confidence in the results of a cost-effectiveness analysis.[5,6]

Recommendations for the extrapolation of survival data[5] published in 2013 state that six standard parametric models (exponential, Weibull, Gompertz, log-logistic, log-normal and generalized gamma) should be fitted to observed data and assessed for goodness of fit (e.g. using visual assessment and statistical measures such as the Akaike information criterion (AIC)) and for the plausibility of the extrapolated tail. The latter can be assessed by comparison with longer-term external data, expert opinion and by considering the biological plausibility of the duration of drug effects on the disease course.[5] Whilst these recommendations were primarily focused on fitting models to RCT data, they are fundamentally about applying a systematic framework for model selection when extrapolating from incomplete survival data. Hence, they are also applicable in an observational data setting. If these models are found to be inappropriate, it is recommended that more flexible and complex modelling methods may be required, such as the flexible parametric spline models developed by Royston and Parmar.[7] The inclusion of knots in spline models increases the models' ability to fit the observed data, however, spline models were originally developed to model observed data with complex hazards, and not for the purpose of extrapolation.[7] Thus there are concerns that while the models may improve fit to the observed data, this may not translate into improvements in the prediction of unobserved data.[8]

Recent studies have compared alternative approaches to extrapolating overall survival from clinical trials of immunotherapies, applying spline, cure and landmark models to reflect the proposed novel mechanism of action for immunotherapies. However, these studies were only able to compare extrapolated survival with observed survival to around three[9] and five[10] years, which provides a limited basis for assessing predictive accuracy.

Observational data, available via cancer registries, can be used to investigate the performance of different extrapolation methods. While there are differences between registry and trial data (e.g. registries have larger cohort sizes with greater heterogeneity and different censoring patterns), the longer follow-up durations in registries enable extrapolated outcomes to be compared to known survival outcomes. The Surveillance, Epidemiology, and End Results (SEER) Program provides an opportunity to investigate the performance of different extrapolation methods for patients with a *de novo* diagnosis of advanced (metastatic) cancer over a long time horizon. SEER is a population-based registry of cancer patients managed by the National Cancer Institute (NCI) in the United States (US).[11] Large cohorts of patients with advanced breast, colorectal, non-small cell lung (NSCLC), small cell lung (SCLC) and pancreatic cancer are available in the SEER registry data. These cancers each have very different survival distributions and hazard shapes, thus providing a diverse set of case studies in which to investigate extrapolation methods.

This research used the SEER data to investigate how well standard parametric models and flexible parametric spline models estimate observed 10-year survival outcomes (restricted mean survival time (RMST) and surviving percentage) when fitted to 15 different registry cohorts with follow-up times right-censored at 20%, 35% and 50% survival. The cohorts include large patient numbers and little censoring during the follow-up period, enabling the predictive accuracy of the survival models to be assessed when fitted to data with little uncertainty. Future analyses should build on this work to examine model fit with more uncertain data (with smaller sample sizes and non-uniform censoring throughout the observed survival period) to determine the relative importance of alternative models' handling of uncertainty in the data and their ability to fit the shape of the observed data.

Methods

SEER data and the patient cohorts

SEER registry data from 1973 to 2015 (the SEER9 grouping) were obtained for breast, colorectal, NSCLC, SCLC and pancreatic cancer patients.[11,12] Patients were included in the analysis cohorts if they were aged 18 years or over at diagnosis and their first diagnosed tumor was advanced or metastatic cancer (defined as 'distant' in the 'SEER historic stage A' variable). Breast, colorectal and pancreatic cancer patients were included if diagnosed between 1973 and 2005 (enabling a minimum potential follow-up time of 10 years), while NSCLC and SCLC patients were included if diagnosed between 1988 and 2005 (as the SEER historic stage A variable was not collected for these patients before 1988). All included breast cancer patients were female. Patients were excluded from the cohorts if they did not have a completed event indicator (i.e. unknown death / censoring status) or were diagnosed with the cancer at autopsy.

Survival times in the SEER data are reported in one month increments, with partial months rounded down to the last whole month. In cohorts with shorter survival times (e.g. pancreatic cancer), this clustering of survival times meant there were a limited number of data points (unique survival times) to which survival models could be fitted. To increase the number of data points and facilitate more robust model fitting, daily survival times were generated for each patient by randomly allocating them a day of the month that was added to their monthly survival time. Consistent with the SEER algorithm for calculating survival times, each month lasted 30.44 days (365.25/12). Patients were excluded from the cohorts if they had a survival time of zero days. Patients with follow-up times greater than 10 years were right-censored at 10 years.

There were significant differences in survival outcomes by age group at diagnosis for all five of the included cancers, with higher ages associated with poorer survival. This enabled each cancer cohort to be split into three separate age groups (18 to 59 years, 60 to 69 years, and 70 years and over) creating 15 analysis cohorts, each with different survival distributions and hazard shapes.

For each of the 15 cohorts, the time-point at which 20%, 35% and 50% of patients were still alive was identified from the Kaplan-Meier estimator. The follow-up times for all surviving patients were right-censored at each of these time-points to generate three case study datasets for each of the 15 cohorts (45 datasets).

Survival models

Six standard parametric models were fitted to each dataset, including the exponential (assumes a constant hazard), Weibull and Gompertz (assume a monotonically increasing or decreasing hazard), the log-logistic and log-normal (allow for non-monotonic hazards), and the more flexible generalized gamma.[6,13]

Flexible parametric spline models were fitted to each dataset as restricted cubic splines – where piecewise cubic polynomials are joined at 'knots' while being constrained to be smooth across the curve and linear beyond the boundary knots.[14] Spline models were fitted on three scales – proportional hazards, proportional odds, and normal / probit. When no knots are specified these models are equivalent to the Weibull, log-logistic and log-normal parametric models, respectively.[15] Beyond the boundary knots where the models are constrained to be linear, the spline hazard behaves like a Weibull, the spline odds behaves like a log-logistic and the spline normal behaves like a log-normal. By default, knot locations are specified based on the number of knots chosen and the percentiles of uncensored event times.[14] Following recommendations, the models on the three scales were all fitted with one to three internal knots; for each scale, the model with the lowest AIC was selected.[7,16]

Analysis

The predicted 10-year RMST and the predicted percentage of patients surviving at 10 years were calculated for all fitted models. To summarize the performance of each model across the 15 cohorts, boxplots were generated to show the median difference between predicted and observed outcomes (RMST and percentage surviving) at each follow-up duration, as well as the interquartile range (IQR) and full range. Differences were calculated as the predicted value minus the observed value. Positive differences indicated the model over-predicted the observed value, while negative values indicated the model under-predicted.

Boxplots were also generated to represent absolute differences in predictions with alternative follow-up times, as described for RMST for datasets with 50% and 35% of patients remaining alive:

$$\text{Absolute difference between model predictions for different follow-up durations} = \text{absolute}\,(P_{50\%} - O) - \text{absolute}\,(P_{35\%} - O),$$

where O is the observed RMST, and $P_{50\%}$ and $P_{35\%}$ are the predicted RMSTs when survival has been right-censored at the time at which 50% and 35% of patients remain alive, respectively. Positive differences indicate predictions from the dataset with the longer follow-up (e.g. 35% surviving in this example) were more accurate.

To compare performance of the alternative fitted curves, the frequency of cases in which models predicted observed RMST within one month and the percentage surviving within 1% were calculated. These cut off points were subjectively selected to reflect decision makers' interpretation of the accuracy of the fitted models.

To mimic the standard extrapolation approach of selecting the 'best' fitting model, the standard parametric model and the spline model with the lowest AIC were selected for each of the 15 cohorts at each follow-up duration. Model fit for the selected parametric and spline models was compared using the boxplots and performance measures described above.

In order to determine whether models performed well in fitting the observed segment of the data or in extrapolating the unobserved tail, the 10-year RMST was calculated for each segment separately,

and compared for each follow-up duration using boxplots of the difference between the predicted and observed values.

Survival and hazard plots are provided in the supplementary materials, with example plots included in the paper. The observed hazard was plotted both as a smoothed hazard using kernel-based methods to smooth out the noise[9,17], and as a piecewise exponential (PE) hazard function with three monthly bin widths (where the hazard has been calculated as the number of events in the bin divided by the total follow-up time in the bin).[17] The PE hazard enables some of the underlying features of the data that are lost in the smoothing manipulations to be seen. The survival and hazard plots were examined by the authors to assess the visual fit of the models to the observed data in relation to the models' predictions.

Software

All analyses were undertaken in R (versions 3.4.1 and 3.5.3). The *flexsurv* library (https://cran.r-project.org/web/packages/flexsurv/index.html) was used to fit the standard parametric and flexible parametric spline models.[15] *Flexsurv* uses maximum likelihood estimation to estimate model parameters. The *muhaz* library (https://cran.r-project.org/web/packages/muhaz) was used to generate the smoothed hazard (*muhaz* function) and the PE hazard function (*pehaz* function).[17]

Funding source

This study was funded by a grant from the National Health and Medical Research Council (NHMRC). Additionally, NRL was supported by the National Institute for Health Research (NIHR Post Doctoral Fellowship, Dr Nicholas Latimer, PDF-2015-08-022) and is now supported by Yorkshire Cancer Research (Award reference number S406NL). The funding agreements ensured the authors' independence in study design, data interpretation, writing and publishing the report. The views expressed in this publication are those of the author(s) and not necessarily those of the NHS, the NIHR, the Department of Health and Social Care, or Yorkshire Cancer Research. RW chairs the Medical Services Advisory Committee for the Commonwealth of Australia and the views in this paper do not reflect those of the Commonwealth.

Results

Cohort characteristics

Table 1 presents the characteristics of the 15 cohorts. All cohorts contained large numbers of patients (minimum n = 5407, SCLC 18-59 years), thus sample size was not a constraint in the analyses. Survival distributions and hazard functions (Figure 1) varied across the cohorts, with breast 18-59 years showing the longest survival times (10-year RMST = 42.0 months, 10-year percentage surviving = 14.67), and pancreatic 70 years and over showing the shortest survival times (10-year RMST = 3.89 months, 10-year percentage surviving = 0.21). Follow-up durations at which right-censoring was applied (Table 1) ranged from 1.72 months (pancreatic 70+ years, 50% surviving) to 81.43 months (breast 18-59 years, 20% surviving).

Model predictions when fitted to 10-year data

Figure 2 summarizes the differences in the accuracy of model predictions for both RMST (Figure 2a) and percentage surviving (Figure 2b) at 10 years and with right-censoring at the timepoints at which

20%, 35% and 50% of the cohort remain alive. The results are generally consistent across the two survival outcomes. The actual predicted values are provided in Appendix A where individual predictions of RMST (Figure A1, Table A1) and percentage surviving (Figure A2, Table A2) are plotted with 95% confidence intervals and tabled for each model, cohort and follow-up duration.

Spline models with three internal knots were most frequently selected on the basis of AIC for both the 10-year and right-censored datasets (Appendix B: Table B1). When fitted to the full 10-year data the three spline models performed better than any of the six standard parametric models (Figure 2), giving predictions of both RMST within one month and percentage surviving within 1% for all 15 cohorts (Table 2). Of the standard parametric models, the Gompertz performed the best (Figure 2; Table 2) giving predictions of both RMST within one month and percentage surviving within 1% for 14 of the 15 cohorts.

Model predictions when fitted to right-censored data

Across the 15 cohorts and three right-censored follow-up durations (45 datasets) the odds scale and normal scale spline models gave a larger number of close predictions than any of the six standard parametric models (Table 2, Figure 2). For the longest follow-up times (20% survival), the log-logistic, log-normal, spline odds and spline normal models produced the most accurate predictions of RMST, but at the shorter follow-up times, the spline odds and spline normal models outperformed the other models. For percentage surviving at 10 years, the log-logistic, log-normal, spline odds and spline normal models performed similarly well over all three follow-up scenarios. When examining the number of cohorts in which the models produced accurate predictions of both the RMST and percentage surviving (Table 2), the spline normal (18/45) and spline odds (17/45) models outperformed the log-logistic (11/45) and log-normal models (7/45).

The exponential and Gompertz were the poorest performing models for both survival outcomes (Table 2, Figure 2), while the Gompertz and generalized gamma had the most variable predictions (Figure 2). In 11/45 datasets (Appendix A: Figure A2) the Gompertz predicted a surviving percentage that plateaued early and high, resulting in highly implausible predictions.

The observed surviving percentages at 10 years were very small (<2.5%) for 9 out of the 15 cohorts (Table 1), and some of the standard parametric models (exponential, Weibull, generalized gamma, and less frequently the Gompertz) tended to predict 0% surviving at 10 years for these cohorts across all follow-up durations. For the five cohorts in which the observed 10-year surviving percentages were less than 1%, these models predicted within 1% of the observed, however, their poor fit to the observed 10-year survival data led to poor predictions of 10-year RMST (Table 2). As an example, for the NSCLC 70+ years cohort censored at 50% surviving (Appendix G2) the standard parametric models fit well to the initial, non-censored portion of the survival curve but rapidly drop to 0% surviving in the extrapolated portion. This drop occurs well before survival in the long-term SEER data decreases to less than 1%, thus giving poor predictions of 10-year RMST.

The generalized gamma was most frequently selected as the ‘best’ fitting parametric model (based on lowest AIC) across the three follow up durations, while the normal scale model was most frequently selected as the ‘best’ fitting spline model (Appendix B: Table B2). The selected spline models (20/45) gave more accurate predictions of both RMST and percentage surviving across the analysis cohorts, outperforming the selected parametric models (7/45) (Figure 2, Table 2).

Examining the predicted RMST separately for the observed portion of the data and the extrapolated tail (Appendix C: Figure C1, Table C1, Table C2) suggests that both the standard parametric and spline models fit well to the observed portion of the data and for all models most of the difference between the predicted and observed 10-year RMST arises in the extrapolated tail. While the spline models show a small advantage over the parametric models in terms of fit to the observed portion of the data, particularly for longer follow-up durations (e.g. 20% surviving), they also more accurately extrapolate the unobserved tails. This indicates that the advantage of the spline models is not solely due to their enhanced fit to the observed part of the data.

Impact of longer follow-up durations on predictions

Figure 3 summarizes the differences in the accuracy of predicted survival outcomes as follow-up durations increase for each of the fitted models. Across the models and cohorts, longer follow-up times tended to improve predictions of RMST and percentage surviving at 10 years, particularly for the standard parametric models. Improvements were generally greater when moving from 50% surviving to 35% surviving, than when moving from 35% surviving to 20% surviving.

The effect of longer follow-up times on the predictions of percentage surviving was less clear than the effect on RMST, predominantly due to the 9 cohorts with very small surviving percentages where some of the standard parametric models (exponential, Weibull, generalized gamma, and less frequently the Gompertz) tended to predict 0% surviving for all follow up durations.

Assessment of hazards and model predictions

Figure 4 provides illustrative examples of model fit to the observed hazard (and survival) functions alongside the predicted 10-year survival outcomes. The examples were selected to illustrate the different hazard shapes in the observed data. Appendices C to G provide the same plots for all 15 cohorts at all follow-up durations.

The shape of the hazard functions varied between cohorts in the first two years of follow-up, with three key shapes discernible. In its simplest form the hazard function decreased across the entire 10 years of data (e.g. Figure 4c Colorectal 60-69 years and Figure 4d Breast 70+ years). For some cohorts the hazard showed an early increase to a peak, followed by a decrease across the remaining 10 years of data, i.e. a unimodal right-skewed distribution (e.g. Figure 4b NSCLC 60-69 years). The most complex hazard shape seen in the study cohorts was an initial decrease, followed by an increase to a peak before decreasing across the remaining 10 years of data (e.g. Figure 4a SCLC 18-59 years), i.e. a bimodal right-skewed distribution.

For all 15 cohorts the hazard showed a long-term decline, and that decline commenced at a time-point within the first two years of follow-up. However, there were large differences in the percentage surviving at that time-point. Predictions tended to be less accurate for cohorts with more complex shapes (bimodal distributions) and shorter censoring times (where the commencement of the long-term decline of the hazard was not captured).

For the SCLC 18-59 years cohort censored at 35% surviving (bimodal right-skewed hazard shape; Figure 4a), the complex shape and the short right-censoring time meant the data were censored before the hazard had commenced its long-term decline. The predicted outcomes show that the exponential, Weibull, spline odds and spline normal models generated the most accurate predictions. However, the plot of the observed and predicted hazards shows that over-time the exponential and

Weibull models over-predicted, under-predicted, and then over-predicted the hazard. However, because these balanced out, their predictions of RMST were accurate.

The spline odds model, and the spline normal model to a lesser extent, fitted well to the observed hazard, despite the hazard not having peaked at the cut-off time-point. Beyond the last event in the censored data, the spline odds behaves like a log-logistic model, predicting a more steeply decreasing hazard than the spline normal model, which has the longer tail associated with a log-normal distribution. The log-logistic and log-normal models were unable to model the early hazard complexity but predicted a decreasing long-term hazard. The hazard functions for the spline hazard, generalized gamma, Gompertz, and Weibull models were misdirected by the increasing hazard at the end of the censored data and predicted increasing long-term hazards.

For the NSCLC 60-69 years cohort censored at 35% surviving (unimodal right-skewed hazard shape; Figure 4b), the spline normal, spline odds, log-logistic and log-normal models produced the most accurate predictions of RMST. These four models tended to slightly underestimate the hazard, whilst the other five models mostly overestimated the observed hazard. The early peak in the observed hazard enabled relatively accurate predictions from the log-logistic and log-normal models, but the increased flexibility of the spline normal and odds models improved the fit to the peak and thus improved their longer-term predictions. However, it is noted that the spline normal model achieves the best prediction because it under- and then over-estimates the hazard (as does the log-normal) – a feature of the longer tail associated with the log-normal model.

For the colorectal 60-69 years cohort censored at 35% surviving (simple hazard shape; Figure 4c), the spline odds and spline normal provided the most accurate predictions of RMST. The flexibility of these models enabled them to fit the flattened hazard shape observed before the censoring point, while still predicting declining long-term hazards similar to the observed hazard. For the breast 70+ years cohort censored at 35% surviving (simple hazard shape with a flat tail; Figure 4d), the Weibull and spline hazard models fitted very well to the observed hazard with its relatively flat decline. In contrast to a cohort where the percentage surviving approached zero early in the follow-up period (e.g. SCLC 18-59 years cohort in Figure 4a), the larger 10-year percentage surviving in this breast cohort (4.51%) meant that model fit in the later portion of the tail had a greater influence on the estimated RMST.

Discussion

Across the 45 analyses in which the models were fitted to right-censored data, the spline odds and spline normal models most frequently predicted accurate estimates of longer term survival. Although based on large datasets with little censoring during the observed follow-up period, these findings may allay concerns that while spline models provide good fit to the observed data, they may be poor extrapolators.[8] The spline models tended to better represent the early complexity of the hazard functions compared to the six standard parametric models. The spline odds and spline normal models assume shapes corresponding to log-logistic and log-normal models beyond the boundary knot, respectively, which tended to better represent the declining hazards in the tails of the observed hazard functions.

Examining the hazard function is a strongly recommended component of assessing model fit.[6,9] In the analyses reported in this paper, the hazards provided a useful adjunct to the survival function, providing insight into the underlying shape of the observed data and the fitted models that enabled us to differentiate between models that appeared to fit the censored survival data equally well but gave very different 10-year predictions. Plotting the observed hazard as both a smoothed hazard function and as a PE hazard using three monthly increments provided valuable insight into both the overall hazard shape (smoothed) and local deviations in the hazard (PE) that influenced model fit to the censored data.

Across the fifteen cohorts, all observed hazards declined over the longer term and that decline commenced within two years of follow-up. Extrapolation may be poor in cases where the hazard is expected to be complex (e.g. uni- or bi-modal right-skewed hazard distributions) and the hazard in the right-censored data has not begun its long-term decline. In such cases follow-up should be continued to further elucidate the shape of the hazard and inform later additional extrapolation analyses. However, the spline odds and spline normal models may be used to inform an interim funding (or pricing) decision, as these models always assumed a decreasing hazard after the last event in the censored data and tended to fit to the tail of the hazard functions. The spline hazard model assumed a longer, flatter tail in the hazard and so may be the more appropriate choice for cancers in which hazards are expected to decline more slowly. The long-term decrease in hazards observed across all cohorts would seemingly support the use of log-logistic or log-normal survival models, however we found that these models predicted decreasing hazards at much earlier time points than the spline odds, spline normal and spline hazard models, and may lead to significant overestimation of RMST if the hazard is not uni-modal with an early peak. The other models failed to adequately reflect the declining hazard, and any accurate predictions were due to over-prediction of the later hazard compensating for under-prediction of the early hazard.

Model choice for data with simpler hazards (where the hazard declines throughout the follow-up duration) should be informed by the expected rate of decline in the long-term hazard. Even for simpler hazards, spline models tended performed better than parametric models as they were able to fit the early part of the data more accurately. Generally the spline odds and spline normal models provided a good fit to this shaped hazard, however the spline hazard may be the more appropriate model choice when a slower decline in the hazard is expected.

The predictions of the standard parametric models were more likely than spline models to improve with longer follow-up durations, with the largest improvements noted for the Gompertz and generalized gamma models. Amongst the spline models, the spline odds generally improved with longer follow-up durations, but the spline normal and spline hazard were less influenced by follow-up duration. This suggests the spline models may be more robust than the standard parametric models at shorter follow-up durations.

Across the fifteen cohorts, the results indicate that spline models are able to extrapolate well, at least in large datasets with little censoring during the observed follow-up. The results suggest that spline models should be included in the standard set of models that are fitted when extrapolating cancer survival data. The choice of spline model should be informed by prior knowledge of the expected shape or rate of decline of the hazard beyond the observed data, for example through the use of relevant external data or expert opinion. The spline hazard assumes the lowest rate of decline and the spline odds assumes the greatest rate of decline. The spline normal is in-between the two,

but tends to be closer to the spline odds than the spline hazard. The analyses of registry data presented in this paper provide a broad indication of long-term hazard functions for different advanced cancers, however the cohorts are heterogeneous. Data for more homogenous patient groups may be required to inform the expected long-term hazard shape for specific trial cohorts. Care should be taken when fitting models with a declining long-term hazard to ensure the model's hazard does not fall below background mortality rates, particularly as hazards will increase as patients age.

Limitations

During the current analyses comparisons have been made between the six standard parametric models and the three spline models. In reality, an economic analysis would aim to select a single 'best' fitting model based on model fit to the observed data and using long-term population-level data and clinical input to inform expectations of long-term survival.[5] The authors experimented with options for selecting the 'best' parametric model and 'best' spline model for each cohort at each follow-up duration to enable a comparison to be made (e.g. each author independently examined plots showing the censored portion of the data and all fitted models extrapolated to 10 years), however without long-term population level data or expert clinical opinion to inform long-term survival there was significant variability in the models selected by different authors. To reduce the subjectivity of model selection, the standard parametric model and the spline model with the lowest AIC was selected for each cohort and used for comparison, however, the authors acknowledge there are important limitations associated with selecting a model for extrapolation based only on the assessment of model fit to the observed data.[5]

Advanced breast, colorectal, SCLC, NSCLC and pancreatic cancer cohorts were selected for the analysis due to the variation in their survival distributions and hazard shapes. Further dividing the cancer cohorts by age increased this variation, creating fifteen analysis cohorts and enabling model fit to be examined a across a diverse range of hazard shapes. The generalizability of the findings to other cancer cohorts will depend on the hazard shape of those cohorts, in particular whether the hazard is expected to decline in the long-term and the rate of that decline.

The SEER registry cohorts used in the analysis are made up of large, highly heterogeneous populations with very little censoring other than that artificially introduced in the analysis. Trial eligibility criteria mean that enrolled patients are frequently younger, healthier, have less comorbidities, and are more homogenous in terms of cancer subtype than cancer patients in the general population.[18,19] Differences between clinical trial and general patient populations will affect survival[18,20,21], and may result in differences in the underlying hazard and the overall shape of the respective survival curves. The use of 15 cohorts tested the fit of the selected survival models to a wide range of underlying hazards and survival functions with known 10-year survival outcomes in which the accuracy of the extrapolation could be evaluated. In all cohorts used in the analysis the observed hazard declined over the long-term, however this may not be the case for other cohorts or cancers.

Patient numbers in clinical trials are much smaller than those seen in the SEER registry cohorts. While a trial may run for a defined length of time, patients are often recruited over a prolonged period, resulting in variable follow-up times for patients who are administratively censored at the end of the trial (rather than having all surviving patients censored with the exact same duration of follow-up, as in the registry analyses). This often results in highly uncertain tails in the survival curves, which could

reduce the accuracy of extrapolations. This uncertainty can be seen when the tail of the survival curve shifts between the interim and final datacuts of a clinical trial. Spline models are likely to follow the tails of the observed data more closely than the standard parametric models. Thus spline models may perform well when fitted to SEER registry cohorts with low levels of uncertainty, but their performance may be more influenced by uncertainty in the tail of trial data. Our analyses provide evidence that the spline models perform well under 'ideal' conditions – where the analysis cohorts were large and there was little right-censoring during the observed follow-up period. Subsequent analyses to further test the performance of spline models could involve repeatedly taking random samples from registry cohorts and introducing censoring throughout the sampled time period (noting that registry-derived cohorts may be more heterogeneous than clinical trial populations). Alternatively, studies have compared extrapolations of interim datacuts of clinical trial data to the results of later datacuts,[9,10] however, to date the duration of follow-up for the later datacuts used in these studies remains relatively short. Further analyses, using data with greater uncertainty, will enable the assessment of whether the handling of uncertainty in the data is more important than the ability of the model to fit the shape of the observed data.

Bias in RMST was used as the primary outcome as it is the measure of most interest to decision makers in health technology assessments. Model performance was summarized using the number of cohorts in which each model predicted within one month of the observed RMST and within 1% of the observed percentage surviving. Other performance measures (e.g. how frequently the observed value was within the predicted confidence intervals) were found to be less informative when comparing models in these analyses but have been included in Appendix A for thoroughness.

The presented analyses used historical registry cohorts with known survival times in order to compare predicted and observed survival outcomes. A 10-year time horizon was chosen for the extrapolation to limit the number of patients excluded because of insufficient follow up durations (e.g. a 20-year time horizon would have excluded patients diagnosed after 1995, while the 10-year time horizon excluded only those diagnosed after 2005). Other time horizon durations could have been chosen, and it is possible that the time horizon may affect the results, for example, prediction differences between models may be small at 10 years but become larger at 20 years. This is more likely for cohorts in which the surviving proportions are relatively large at 10 years.

In recent clinical trials, patients with advanced cancers receiving newer cancer therapies (e.g. immunotherapies) have been shown to have distinctly different survival functions to similar patients receiving traditional chemotherapies.[22] However, the presented analyses using historical registry data remain relevant as older chemotherapies are commonly used as the comparators for newer treatments.

For immunotherapy survival data, studies have compared extrapolations from earlier datacuts to survival curves and outcomes observed at later datacuts.[9,10] Ouwens et al[9] extrapolated data with up to two years follow-up and compared predictions to observed survival data with up to three years follow-up. The fitted spline models performed as well as any other models. Bullement et al[10] compared models fitted to data with up to four years follow-up to survival observed in data with up to six years follow-up. Survival predictions were also compared to survival in registry data at 10 and 15 years, with an assumption that predictions should be no worse than registry survival. This study reported a noticeable improvement in survival predictions for the immunotherapy arm at the later datacut using cure models.

Cure models reflect an assumption that the population of interest is comprised of groups with different mortality hazards, i.e. cured and non-cured patients. Cured patients are assumed to experience age-specific general population mortality hazards. An issue with cure models is the variability in the cured proportions that are fitted by different models. In the cure models fitted by Ouwens et al[9] the cured proportions ranged from 0% to 23%, which resulted in lifetime predicted survival varying from 1.75 years to 5.81 years. More case studies evaluating the accuracy of alternative extrapolation models for patient cohorts receiving immunotherapies, with longer follow-up, are required to confirm the advantage of cure models when extrapolating immunotherapy survival. Cure models are not reported in the current paper because the concept of cure is less relevant to historical registry cohorts, but also because there is no consensus on the application of cure models and the exploration of alternative approaches to the application of cure models is beyond the scope of this paper.

Recent analyses by Gibson et al[23] suggests that spline models are able to adequately model the long-term plateau seen in progression free survival for a proportion of immunotherapy recipients. The observed patterns of the longer-term hazards predicted by the spline normal, odds and hazard models may provide decision makers with a less complicated and adequately robust process for assessing longer-term survival benefits. Further improvements in the performance of spline models may be observed through the application of the relative survival framework that reflects the potential for cure by incorporating general population mortality hazards.[24]

Conclusions

Fifteen large population-based cohorts of patients with *de novo* diagnoses of advanced breast, colorectal, SCLC, NSCLC and pancreatic cancer were used to evaluate predictions of long-term survival outcomes from six standard parametric and three flexible parametric spline models. In the analyzed cohorts, where there was little censoring during the observed follow-up period, the spline models generated the most accurate predictions across a range of hazard functions. Inspection of the hazard functions informed generalizable recommendations that may inform the selection of survival models when extrapolating cancer survival data. The three flexible parametric spline models should be routinely included in the set of fitted survival models when extrapolating cancer survival data. Replicating these analyses in clinical trial data for a variety of cancers with different survival distributions would provide further insight into the generalizability of the results.

Conflicts of Interests

Dr Michael Sorich reports grants from Pfizer outside the submitted work. Dr Nicholas Latimer reports consultancy fees from Astra Zeneca, Bluebird Bio, Merck EMD Serono, Pierre Fabre outside the submitted work outside the submitted work. The other authors declare that there are no conflicts of interest.

Supplementary materials

Appendix A. Supplementary figures and tables of predicted survival outcomes

- Figure A1. Difference in 10−year restricted mean survival time (RMST) by model, cohort and follow−up duration
- Figure A2. Difference in percentage surviving at 10 years by model, cohort and follow−up duration
- Table A1. Difference in 10-year restricted mean survival time (RMST) by cohort, model and follow−up duration (months)
- Table A2. Difference in percentage surviving at 10 years by cohort, model and follow−up duration (%)

Appendix B. Supplementary table of spline model knots

- Table B1. Number of internal knots selected for each spline model
- Table B2. Models selected based on lowest AIC

Appendix C. Supplementary table of spline model knots

- Figure C1. Model predictions of restricted mean survival time (RMST) across the 15 cohorts for (a) the observed portion of the data and (b) the extrapolated (unobserved) tail of the data
- Table C1. Difference in restricted mean survival time (RMST) for the observed portion of the data by cohort, model and follow-up duration (months)
- Table C2. Difference in restricted mean survival time (RMST) for the extrapolated (unobserved) tail of the data by cohort, model and follow-up duration (months)

Appendices D to H. Supplementary figures of survival and hazard functions, model fit and predicted survival outcomes for all cohorts at all follow-up durations

- Appendix D: Breast cancer cohorts
- Appendix E: Colorectal cancer cohorts
- Appendix F: SCLC cohorts
- Appendix G: NSCLC cohorts
- Appendix H: Pancreatic cancer cohorts

References

1. National Institute for Health and Care Excellence (NICE). Guide to the methods of technology appraisal 2013. Available at https://www.nice.org.uk/process/pmg9. Accessed on 11 July 2018.
2. Department of Health. Guidelines for preparing a submission to the Pharmaceutical Benefits Advisory Committee, Version 5.0, September 2016. Canberra: Commonwealth of Australia; 2016.
3. Bagust A, Beale S. Survival Analysis and Extrapolation Modeling of Time-to-Event Clinical Trial Data for Economic Evaluation. Medical Decision Making. 2014;34(3):343-51.
4. Connock M, Hyde C, Moore D. Cautions Regarding the Fitting and Interpretation of Survival Curves. PharmacoEconomics. 2011;29(10):827-37.
5. Latimer NR. Survival Analysis for Economic Evaluations Alongside Clinical Trials—Extrapolation with Patient-Level Data. Medical Decision Making. 2013;33(6):743-54.
6. Latimer N. NICE DSU Technical Support Document 14: Undertaking survival analysis for economic evaluations alongside clinical trials-extrapolation with patient-level data. Sheffield, UK: Decision Support Unit, ScHARR, University of Sheffield; 2011.
7. 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):2175-97.
8. Tremblay G, Livings C, Crowe L, Kapetanakis V, Briggs A. Determination of the most appropriate method for extrapolating overall survival data from a placebo-controlled clinical trial of lenvatinib for progressive, radioiodine-refractory differentiated thyroid cancer. Clinicoecon Outcomes Res. 2016;8:323-33.
9. Ouwens MJNM, Mukhopadhyay P, Zhang Y, Huang M, Latimer N, Briggs AJP. Estimating Lifetime Benefits Associated with Immuno-Oncology Therapies: Challenges and Approaches for Overall Survival Extrapolations. 2019.
10. Bullement A, Latimer NR, Bell Gorrod H. Survival Extrapolation in Cancer Immunotherapy: A Validation-Based Case Study. Value in Health. 2018.
11. National Cancer Institute (NIH). SEER Research Data Record Description: Cases diagnosed in 1973-2015*. Bethesda, Maryland, USA: National Cancer Institute: Surveillance, Epidemiology, and End Results Program; 2018.
12. National Cancer Institute, DCCPS, Surveillance Research Program, Surveillance Systems Branch. Surveillance, Epidemiology, and End Results (SEER) Program (www.seer.cancer.gov) Research Data (1973-2015). Released April 2018, based on the November 2017 submission. 2018.
13. Collett D. Modelling survival data in medical research. 3rd ed. Boca Raton: CRC Press, Taylor & Francis Group; 2015.
14. Lambert PC, Royston P. Further development of flexible parametric models for survival analysis. Stata Journal. 2009;09(2).
15. Jackson CH. flexsurv: A Platform for Parametric Survival Modeling in R. Journal of statistical software. 2016;70(1).
16. Royston P. Flexible parametric alternatives to the Cox model, and more. Stata Journal. 2001;1(1).
17. CRAN. muhaz: Hazard function estimation in survival analysis. Available at https://cran.r-project.org/web/packages/muhaz/muhaz.pdf
18. Elting LS, Cooksley C, Bekele BN, Frumovitz M, Avritscher EB, Sun C, et al. Generalizability of cancer clinical trial results: prognostic differences between participants and nonparticipants. Cancer. 2006;106(11):2452-8.
19. Kennedy-Martin T, Curtis S, Faries D, Robinson S, Johnston J. A literature review on the representativeness of randomized controlled trial samples and implications for the external validity of trial results. Trials. 2015;16:495.
20. Lobbezoo DJ, van Kampen RJ, Voogd AC, Dercksen MW, van den Berkmortel F, Smilde TJ, et al. Prognosis of metastatic breast cancer subtypes: the hormone receptor/HER2-positive subtype is associated with the most favorable outcome. Breast Cancer Res Treat. 2013;141(3):507-14.

21. Press DJ, Miller ME, Liederbach E, Yao K, Huo D. De novo metastasis in breast cancer: occurrence and overall survival stratified by molecular subtype. Clinical & Experimental Metastasis. 2017;34(8):457-65.
22. Chen T-T. Statistical issues and challenges in immuno-oncology.(Report). Journal for ImmunoTherapy of Cancer. 2013;1(1).
23. Gibson E, Koblbauer I, Begum N, Dranitsaris G, Liew D, McEwan P, et al. Modelling the Survival Outcomes of Immuno-Oncology Drugs in Economic Evaluations: A Systematic Approach to Data Analysis and Extrapolation. PharmacoEconomics. 2017;35(12):1257-70.
24. Andersson T, Dickman P, Eloranta S, Lambe M, Lambert P. Estimating the loss in expectation of life due to cancer using flexible parametric survival models. Stat Med. 2013;32(30):5286-300.

Table 1. Characteristics of the 15 cohorts

Cohort			Observed 10-year outcomes		Time-points of right-censoring (months)		
Cancer	Age group (years)	n	RMST (months)	Percentage surviving (%)	20% surviving	35% surviving	50% surviving
Breast	18-59	10480	42.00	14.67	81.43	42.00	26.03
Breast	60-69	6308	33.96	8.50	59.33	33.46	20.39
Breast	70+	8587	26.00	4.51	44.56	23.99	12.59
Colorectal	18-59	16324	23.53	6.51	31.46	18.66	12.26
Colorectal	60-69	17378	18.59	4.01	25.07	15.13	9.53
Colorectal	70+	32022	13.10	2.05	17.76	9.53	5.26
SCLC	18-59	5407	12.87	2.09	15.69	11.36	8.53
SCLC	60-69	7637	10.38	0.91	13.59	9.76	6.76
SCLC	70+	8933	6.78	0.32	10.49	6.46	3.26
NSCLC	18-59	19414	13.08	2.56	15.59	9.36	6.03
NSCLC	60-69	21796	10.55	1.30	13.20	7.82	4.85
NSCLC	70+	29257	7.88	0.50	10.23	5.76	3.43
Pancreatic	18-59	8710	8.67	1.40	9.46	5.59	3.62
Pancreatic	60-69	9888	5.65	0.43	7.00	4.30	2.72
Pancreatic	70+	15892	3.89	0.21	4.72	2.72	1.72

RMST: restricted mean survival time. SCLC: small cell lung cancer. NSCLC: non-small cell lung cancer. 70+: 70 and over.

Table 2. Number of cohorts in which model predictions were within 1 month of the observed restricted mean survival time (RMST) and 1% of the observed percentage surviving

	Predictions within 1 month of observed 10-year RMST					**Predictions within 1% of observed 10-year percentage surviving**					**Predictions within both 1 month of observed RMST and 1% of observed percentage surviving**				
	10 years	*Right-censored at … surviving* *20%*	*35%*	*50%*	*Total right censored*	*10 years*	*Right-censored at … surviving* *20%*	*35%*	*50%*	*Total right censored*	*10 years*	Right-censored at … surviving *20%*	*35%*	*50%*	*Total right censored*
(Total cohorts)	*(15)*	*(15)*	*(15)*	*(15)*	*(45)*	*(15)*	*(15)*	*(15)*	*(15)*	*(45)*	*(15)*	*(15)*	*(15)*	*(15)*	*(45)*
Exponential	12	2	0	1	3	5	5	5	5	15	5	1	0	0	1
Weibull	13	4	4	2	10	7	6	6	6	18	7	2	3	1	6
Gompertz	14	5	0	1	6	15	4	3	3	10	14	2	0	0	2
Log-logistic	8	7	3	2	12	10	9	8	7	24	7	6	3	2	11
Log-normal	9	7	3	1	11	10	9	6	6	21	6	4	2	1	7
Generalized gamma	14	5	3	1	9	10	6	6	4	16	10	3	3	0	6
Spline hazard	15	5	1	2	8	15	6	6	5	17	15	3	1	1	5
Spline odds	15	8	9	5	22	15	10	7	6	23	15	8	6	3	17
Spline normal	15	8	7	6	21	15	8	7	5	20	15	7	7	4	18
Selected parametric[1]	15	5	3	1	9	15	7	6	5	18	15	4	3	0	7
Selected spline[2]	15	8	7	8	23	15	8	8	6	22	15	7	8	5	20

RMST: restricted mean survival time. [1] The parametric model with the lowest AIC was selected for each cohort. [2] The spline model with the lowest AIC was selected for each cohort

Figure caption – Figure 1

Title

Figure 1. (a) Survival functions, (b) hazard functions and (c) numbers at risk for the 15 cohorts

Caption

Horizontal grey lines on the survival plot (a) show points at which 20%, 35% and 50% of patients survive.

Figure caption – Figure 2

Title

Figure 2. Model predictions of 10-year (a) restricted mean survival time (RMST) and (b) percentage surviving across the 15 cohorts

Caption

Boxplots show the range (whiskers), interquartile range (box) and median (black line within box) of the difference between the predicted and observed outcomes across the 15 cohorts. Models were fitted to the full 10-year data and to the data right-censored at three follow-up durations. Time-points for right-censoring were based on the time-point in each cohort at which the proportion of patients surviving reached 20%, 35% and 50%.

Figure caption – Figure 3

Title

Figure 3. Difference in the accuracy of model predictions with increasing follow-up durations

Caption

Boxplots show the range (whiskers), interquartile range (box) and median (black line within box) of the difference in predictions between the two compared right-censored follow-up durations across the 15 cohorts ((a) compares 50% surviving with 35% surviving, (b) compares 35% surviving with 20% surviving).[1] For each model in each cohort, the difference was calculated as the absolute difference between the predicted and observed for the shorter follow-up time minus the absolute difference between the predicted and observed for the longer follow-up time.

Figure caption – Figure 4

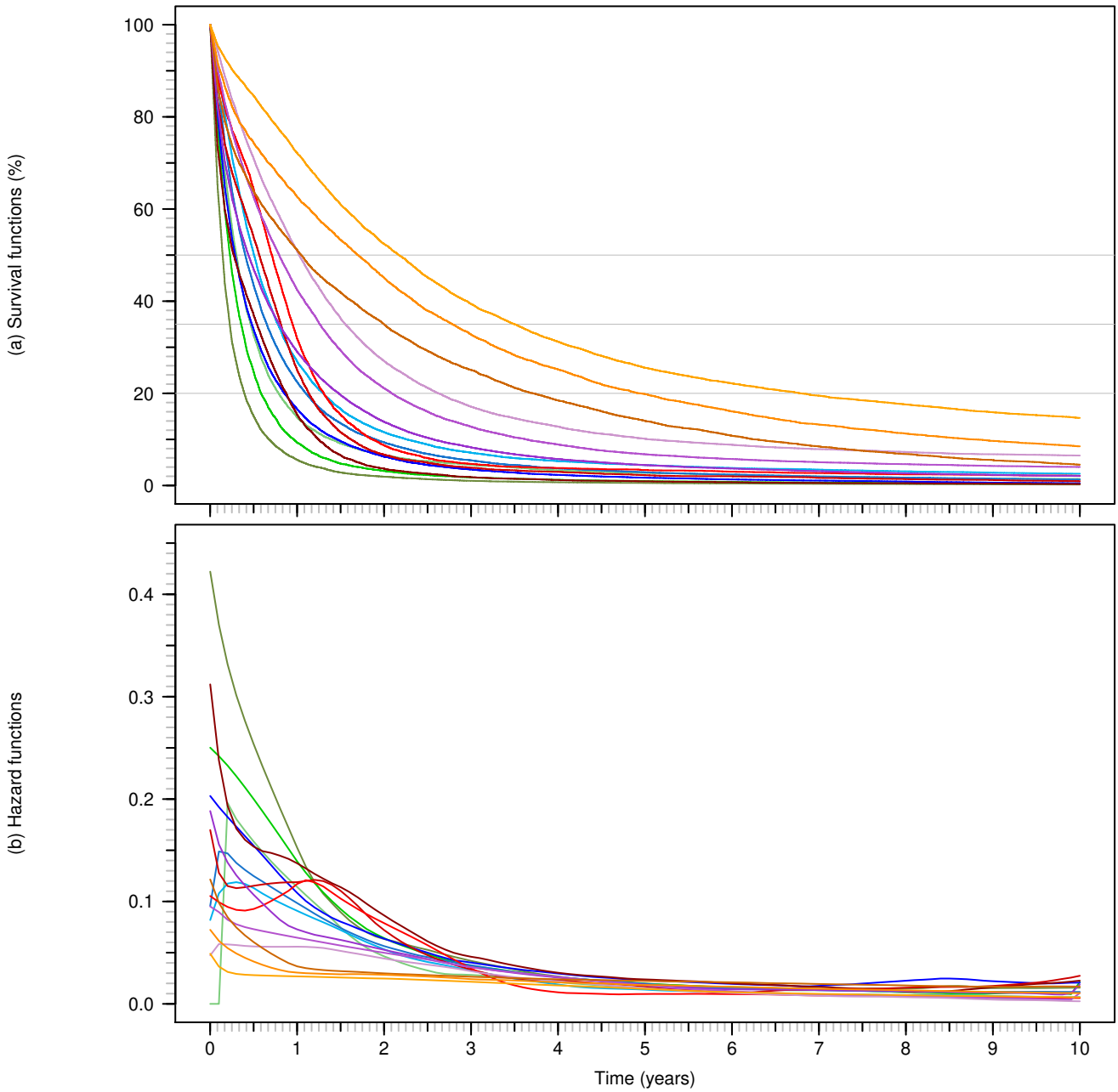
Title

Figure 4. Hazard shape, model fit and predicted survival outcomes for four example cohorts right-censored at 35% surviving

Caption

Plots show the survival and hazard functions for the observed data and fitted models, and the difference between the 10-year predicted and 10-year observed RMST and percentage surviving for each model. The selected examples demonstrate model fit to (a) a complex hazard with a bimodal right-skewed distribution, (b) a hazard with a unimodal right-skewed distribution, (c) a simple declining hazard, and (d) a simple declining hazard with a flatter tail. The time-point at which the percentage surviving reaches 5% is indicated as predictions of the hazard are likely to have the greatest effect on the estimated RMST when the percentage surviving is higher i.e. before this time-point. KM: Kaplan-Meier, PE: piecewise exponential

Figure 1



(c) Number at risk

Breast 18–59 yrs	10480	8857	7550	6375	5475	4710	4103	3645	3247	2923	2651	2462	2296	2153	2015	1912	1815	1723	1633	1568	1501
Breast 60–69 yrs	6308	4686	3946	3354	2837	2386	2059	1778	1579	1379	1242	1117	1006	906	828	765	704	652	606	569	531
Breast 70+ yrs	8587	5481	4389	3590	3000	2502	2150	1816	1583	1377	1201	1059	922	812	725	644	583	517	470	433	384
Colorectal 18–59 yrs	16324	11564	8280	5902	4393	3434	2770	2361	2064	1813	1631	1510	1416	1325	1261	1218	1159	1103	1077	1055	1028
Colorectal 60–69 yrs	17378	10878	7368	5089	3657	2779	2210	1807	1540	1316	1172	1073	993	922	870	834	798	768	729	707	688
Colorectal 70+ yrs	32022	15033	9285	6289	4420	3307	2642	2167	1834	1603	1412	1272	1170	1083	987	905	838	781	727	676	641
SCLC 18–59 yrs	5407	3480	1722	831	469	320	253	219	204	191	178	171	158	150	142	136	127	125	118	114	111
SCLC 60–69 yrs	7637	4115	1910	880	510	377	292	248	223	193	169	153	151	142	125	114	104	95	89	77	68
SCLC 70+ yrs	8933	3312	1377	588	326	228	161	129	107	92	80	69	61	53	47	43	41	36	35	29	26
NSCLC 18–59 yrs	19414	9754	5212	3216	2231	1709	1367	1140	1015	915	844	778	721	682	647	611	577	549	522	498	478
NSCLC 60–69 yrs	21796	9464	4885	2934	2031	1482	1177	955	798	702	613	552	508	458	416	390	363	335	312	291	269
NSCLC 70+ yrs	29257	9899	4840	2813	1824	1298	1012	812	667	564	487	439	377	339	308	278	244	208	173	156	138
Pancreatic 18–59 yrs	8710	2835	1284	794	574	461	388	337	285	254	235	217	197	180	164	153	135	131	125	122	118
Pancreatic 60–69 yrs	9888	2436	918	464	308	222	170	143	120	101	92	82	75	67	61	56	51	48	46	41	39
Pancreatic 70+ yrs	15892	2396	880	449	301	216	160	128	110	95	86	76	69	60	56	52	48	43	37	36	32

Figure 2

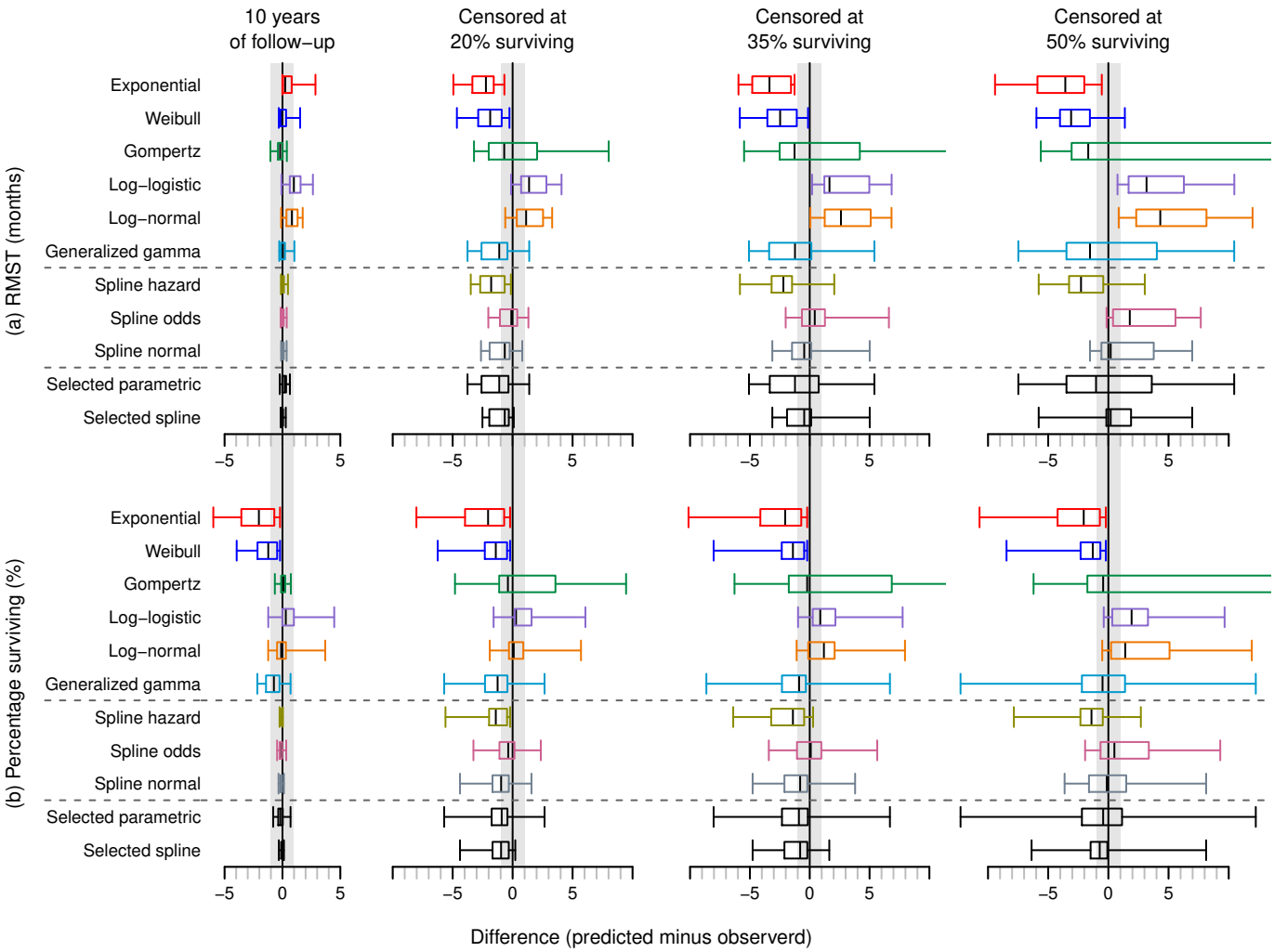
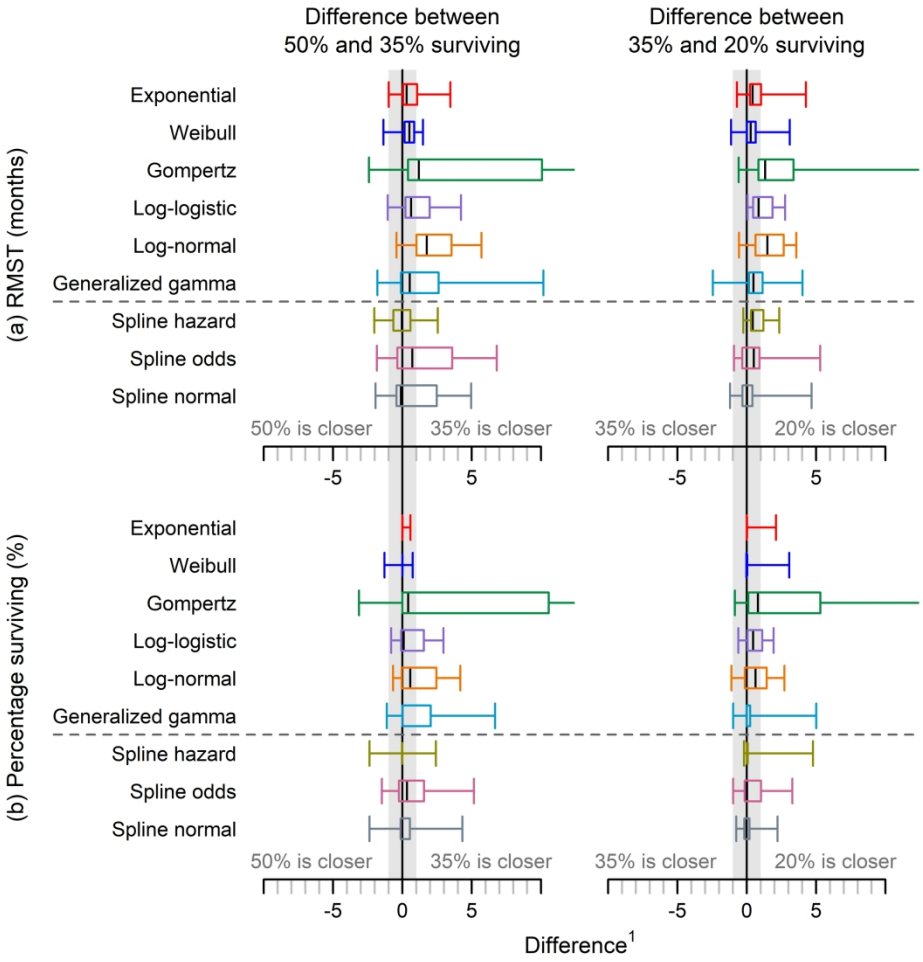
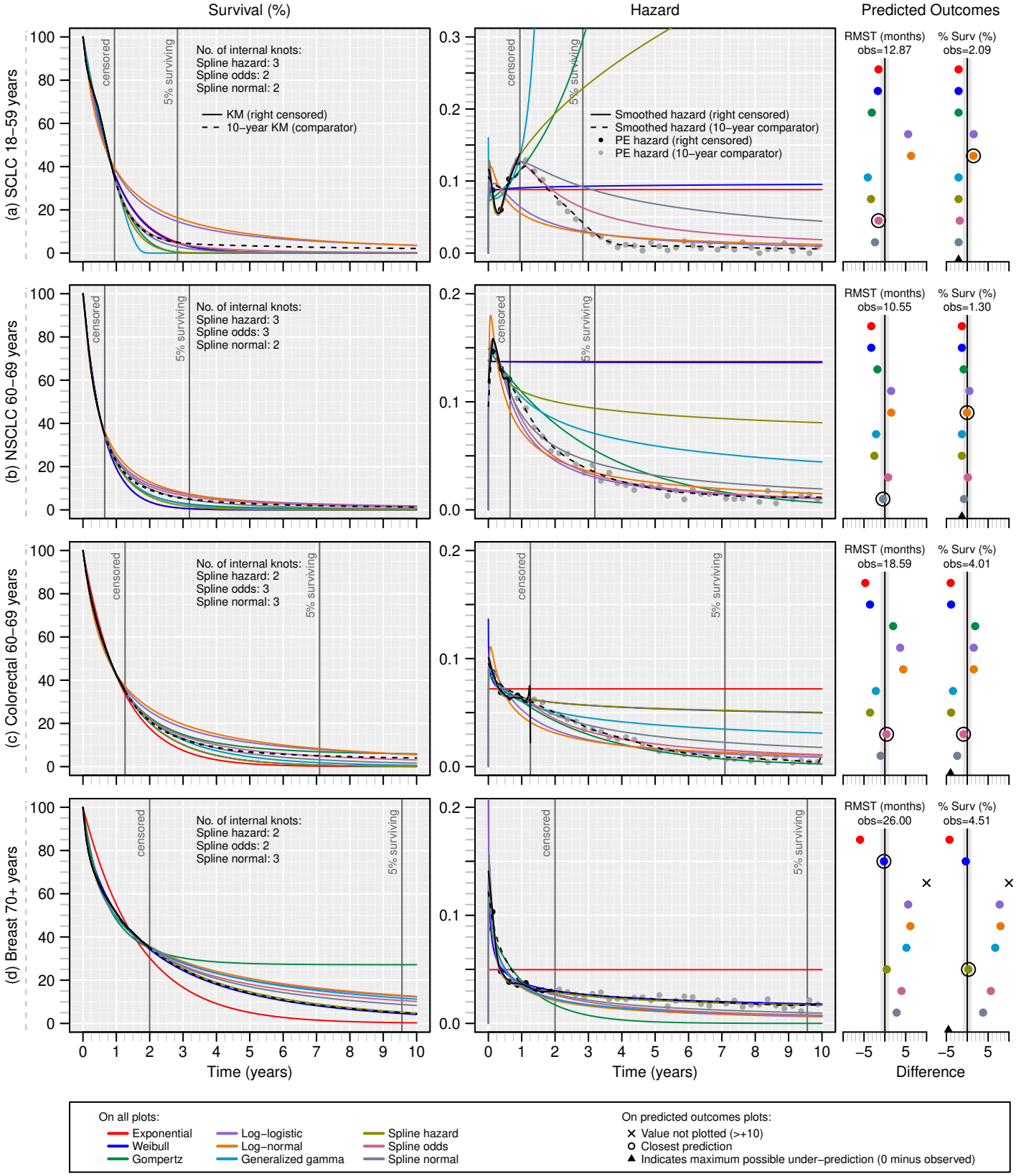

10 years of follow−up
Censored at 20% surviving
Censored at 35% surviving
Censored at 50% surviving
(a) RMST (months)
(b) Percentage surviving (%)
Exponential
Weibull
Gompertz
Log−logistic
Log−normal
Generalized gamma
Spline hazard
Spline odds
Spline normal
Selected parametric
Selected spline
−5
0
5
Difference (predicted minus observerd)

Figure 3



123x127mm (600 x 600 DPI)

Figure 4


Survival (%)
Hazard
Predicted Outcomes
(a) SCLC 18–59 years
(b) NSCLC 60–69 years
(c) Colorectal 60–69 years
(d) Breast 70+ years
censored
5% surviving
No. of internal knots:
Spline hazard: 3
Spline odds: 2
Spline normal: 2
KM (right censored)
10–year KM (comparator)
Smoothed hazard (right censored)
Smoothed hazard (10–year comparator)
PE hazard (right censored)
PE hazard (10–year comparator)
RMST (months)
obs=12.87
% Surv (%)
obs=2.09
No. of internal knots:
Spline hazard: 3
Spline odds: 3
Spline normal: 2
RMST (months)
obs=10.55
% Surv (%)
obs=1.30
No. of internal knots:
Spline hazard: 2
Spline odds: 3
Spline normal: 3
RMST (months)
obs=18.59
% Surv (%)
obs=4.01
No. of internal knots:
Spline hazard: 2
Spline odds: 2
Spline normal: 3
RMST (months)
obs=26.00
% Surv (%)
obs=4.51
Time (years)
Time (years)
Difference
On all plots:
Exponential
Weibull
Gompertz
Log–logistic
Log–normal
Generalized gamma
Spline hazard
Spline odds
Spline normal
On predicted outcomes plots:
Value not plotted (>+10)
Closest prediction
Indicates maximum possible under–prediction (0 minus observed)

Appendices

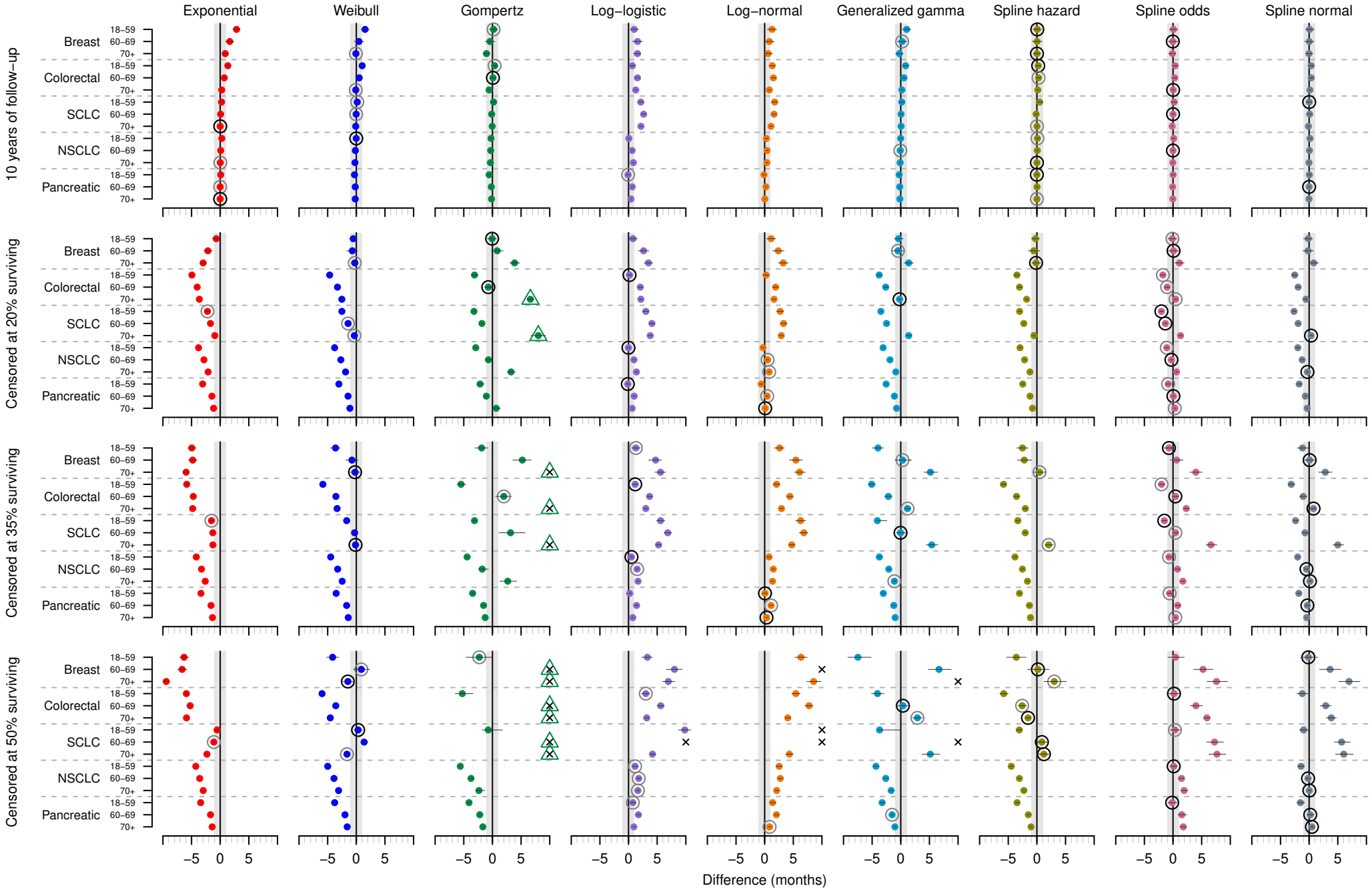


Figure A1. Difference in **10−year restricted mean survival time (RMST)** by model, cohort and follow−up duration

Models were fitted to each cohort with 10 years of follow−up and with the data right censored at the times where 20%, 35% and 50% of patients survived. Differences were calculated as predicted minus observed. Plot shows point estimates with 95% confidence intervals.

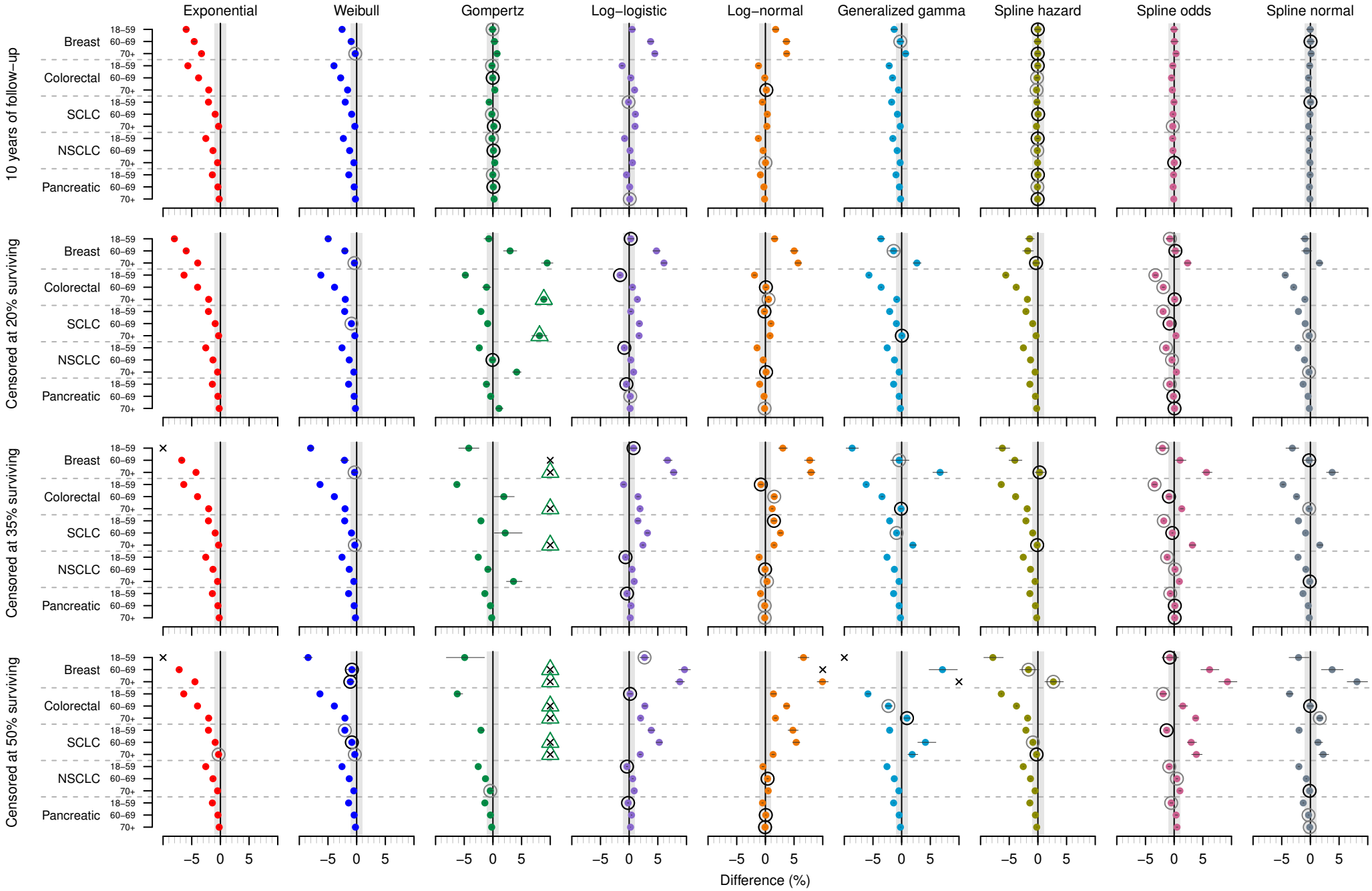


Figure A2. Difference in **percentage surviving at 10 years** by model, cohort and follow−up duration

Models were fitted to each cohort with 10 years of follow−up and with the data right censored at the times where 20%, 35% and 50% of patients survived. Differences were calculated as predicted minus observed. Plot shows point estimates with 95% confidence intervals.

Table A1. Difference in **10-year restricted mean survival time (RMST)** by cohort, model and follow-up duration (months)

		Cohort															Within 1 month (count)	Within CIs[1] (count)	Closest to observed (count)	Closest selected model[2] (count)
		Breast			Colorectal			SCLC			NSCLC			Pancreatic						
		18-59	60-69	70+	18-59	60-69	70+	18-59	60-69	70+	18-59	60-69	70+	18-59	60-69	70+				
Observed RMST (months)		42.00	33.96	26.00	23.53	18.59	13.10	12.87	10.38	6.78	13.08	10.55	7.88	8.67	5.65	3.89				
10 years of follow-up																				
Model	Exponential	2.85	1.66	0.88	1.33	0.71	0.25	0.24	0.08	**0.01**	0.28	0.09	0.02	0.09	-0.01	**0.00**	12	8	2	
	Weibull	1.53	0.46	-0.07	1.03	0.50	-0.09	0.16	-0.02	-0.10	**-0.01**	-0.17	-0.23	-0.32	-0.19	-0.19	13	8	1	
	Gompertz	0.20	-0.43	-1.04	0.38	**0.09**	-0.59	0.20	-0.14	-0.05	-0.25	-0.31	-0.35	-0.58	-0.20	-0.17	14	9	1	
	Log-logistic	0.99	1.57	1.56	0.67	1.56	1.26	2.14	2.63	2.17	0.06	0.63	0.84	-0.09	0.64	0.43	8	2	0	
	Log-normal	1.30	0.82	0.62	1.32	1.53	0.82	1.75	1.64	1.11	0.30	0.42	0.39	-0.11	0.23	0.03	9	4	0	
	Generalized gamma	1.03	0.24	-0.24	0.85	0.56	0.19	0.18	0.05	0.05	-0.02	-0.08	-0.11	-0.28	-0.18	-0.16	14	11	0	
	Spline hazard	**0.07**	0.03	**0.01**	**0.23**	0.29	0.16	0.48	-0.13	0.02	0.10	0.07	**0.00**	**-0.01**	0.03	0.00	15	15	5	
	Spline odds	0.09	**-0.03**	-0.10	0.37	0.30	**0.02**	0.19	**0.01**	-0.15	0.13	**-0.01**	-0.04	0.02	-0.06	-0.04	15	15	4	
	Spline normal	0.10	0.03	-0.02	0.35	0.36	0.13	**0.01**	0.09	-0.12	0.16	0.06	-0.02	0.04	**0.00**	-0.02	15	15	2	
	Selected parametric[3]	0.20	0.24	-0.24	0.38	**0.09**	0.19	0.20	-0.14	0.05	0.06	0.63	-0.11	-0.09	0.64	0.43	15	12	1	3
	Selected spline[4]	**0.07**	0.03	-0.02	**0.23**	0.29	**0.02**	**0.01**	-0.13	-0.15	0.10	**-0.01**	-0.04	**-0.01**	0.03	-0.02	15	15	6	12
Within 1 month (count of 9 models)		5	7	7	6	7	8	7	7	7	9	9	9	9	9	9				
Within CIs[1] (count of 9 models)		4	7	6	4	4	5	7	7	7	7	6	5	7	6	5				
Censored at 20% surviving																				
Model	Exponential	-0.69	-2.15	-2.99	-4.94	-4.02	-3.64	-2.23	-1.70	-0.94	-3.78	-2.85	-2.11	-3.07	-1.45	-1.16	2	1	0	
	Weibull	-0.53	-0.70	-0.27	-4.65	-3.28	-2.53	-2.52	-1.44	-0.34	-3.78	-2.70	-1.87	-3.04	-1.44	-1.12	4	3	0	
	Gompertz	**-0.06**	0.83	3.87	-3.12	**-0.72**	6.63	-3.22	-1.83	8.01	-2.91	-0.70	3.25	-2.15	-1.06	0.65	5	3	2	
	Log-logistic	0.78	2.62	3.47	**0.17**	2.05	2.15	3.01	4.08	3.78	**-0.02**	0.95	1.37	**-0.13**	0.98	0.64	7	4	3	
	Log-normal	1.13	2.37	3.23	0.25	1.92	1.64	2.70	3.30	2.90	-0.29	0.51	0.79	-0.61	0.44	**0.09**	7	3	1	
	Generalized gamma	-0.38	-0.49	1.38	-3.76	-2.63	**-0.21**	-3.47	-2.49	1.38	-3.09	-1.89	-0.87	-2.56	-1.12	-0.73	5	3	1	
	Spline hazard	-0.24	-0.55	**-0.14**	-3.49	-3.03	-1.79	-3.02	-2.29	-0.52	-2.94	-2.15	-1.21	-2.47	-1.19	-0.77	5	3	1	
	Spline odds	-0.11	**0.09**	1.12	-1.76	-1.03	0.45	**-2.03**	**-1.34**	1.31	-1.08	**-0.27**	0.62	-0.85	**0.07**	0.32	8	4	5	
	Spline normal	-0.15	-0.19	0.80	-2.52	-1.92	-0.56	-2.64	-1.92	**0.33**	-1.97	-1.22	**-0.29**	-1.74	-0.66	-0.34	8	3	2	
	Selected parametric[3]	**-0.06**	-0.49	1.38	-3.76	-2.63	**-0.21**	-3.47	-2.49	1.38	-3.09	-1.89	-0.87	-2.56	-1.12	-0.73	5	3	2	2
	Selected spline[4]	-0.24	**0.09**	**-0.14**	-2.52	-1.92	-0.56	**-2.03**	-2.29	-0.52	-1.97	-1.22	**-0.29**	-1.74	-0.66	-0.34	8	3	4	13
Within 1 month (count of 9 models)		8	6	3	2	1	3	0	0	4	2	4	4	3	4	7				
Within CIs[1] (count of 9 models)		8	6	3	2	0	1	0	0	0	2	1	0	1	1	2				
Censored at 35% surviving																				
Model	Exponential	-4.96	-4.80	-5.95	-5.84	-4.70	-4.81	-1.52	-1.29	-1.27	-4.17	-3.27	-2.61	-3.36	-1.60	-1.35	0	0	0	
	Weibull	-3.63	-0.76	**-0.22**	-5.84	-3.56	-3.36	-1.68	-0.31	**-0.14**	-4.48	-3.26	-2.47	-3.52	-1.70	-1.41	4	4	2	
	Gompertz	-1.89	5.22	14.01	-5.48	1.98	15.73	-3.13	3.16	24.54	-4.42	-1.78	2.67	-3.46	-1.55	-1.27	0	0	0	
	Log-logistic	1.28	4.73	5.58	**1.16**	3.67	3.01	5.59	6.85	5.23	**0.51**	1.50	1.65	0.20	1.41	0.74	3	1	2	
	Log-normal	2.61	5.45	6.10	2.07	4.39	2.94	6.27	6.84	4.74	0.77	1.52	1.38	**0.05**	1.13	**0.32**	3	1	2	
	Generalized gamma	-3.98	0.32	5.16	-5.08	-2.16	1.18	-4.10	**-0.04**	5.40	-3.74	-2.10	-1.12	-3.05	-1.23	-0.98	3	2	1	
	Spline hazard	-2.52	-2.20	0.47	-5.83	-3.55	-2.01	-3.33	-2.04	2.05	-3.82	-2.53	-1.65	-3.04	-1.30	-1.13	1	1	0	
	Spline odds	**-0.72**	0.64	3.99	-2.01	**0.41**	2.31	**-1.52**	0.42	6.62	-0.70	0.79	1.72	-0.59	0.82	0.40	9	5	3	
	Spline normal	-1.16	**0.09**	2.82	-3.12	-1.03	**0.77**	-2.37	-0.72	5.01	-2.00	**-0.46**	**0.15**	-1.80	**-0.31**	-0.42	7	3	5	
	Selected parametric[3]	-3.63	0.32	5.16	-5.08	-2.16	1.18	-4.10	-0.31	5.40	-3.74	-2.10	1.65	-3.05	-1.23	-0.98	3	2	0	1
	Selected spline[4]	-1.16	**0.09**	0.47	-3.12	-1.03	**0.77**	-2.37	-2.04	5.01	-2.00	**-0.46**	**0.15**	-1.80	**-0.31**	-0.42	7	4	5	14
Within 1 month (count of 9 models)		1	4	2	0	1	1	0	4	1	3	2	1	3	2	5				
Within CIs[1] (count of 9 models)		1	4	2	0	1	0	0	3	1	0	0	1	3	1	0				
Censored at 50% surviving																				
Model	Exponential	-6.31	-6.65	-9.42	-5.90	-5.24	-5.88	-0.55	-1.11	-2.32	-4.26	-3.58	-2.97	-3.41	-1.69	-1.41	1	0	0	
	Weibull	-4.14	0.88	**-1.48**	-5.98	-3.58	-4.54	**0.31**	1.37	-1.63	-4.99	-3.90	-3.10	-3.78	-1.98	-1.59	2	2	2	
	Gompertz	-2.30	17.84	28.05	-5.26	11.35	26.51	-0.73	30.52	32.14	-5.61	-3.75	-2.36	-4.09	-2.21	-1.67	1	2	0	
	Log-logistic	3.29	8.02	6.94	3.01	5.61	3.17	9.82	10.46	4.18	1.14	1.77	1.63	0.76	1.73	0.93	2	0	0	
	Log-normal	6.34	10.05	8.55	5.44	7.75	4.04	11.98	11.64	4.31	2.53	2.74	2.10	1.38	2.05	0.86	1	0	0	
	Generalized gamma	-7.48	6.67	10.46	-4.07	**0.36**	2.90	-3.70	10.22	5.15	-4.34	-2.63	-1.68	-3.26	-1.52	-1.03	1	1	1	
	Spline hazard	-3.59	**0.18**	3.03	-5.79	-2.58	**-1.55**	-3.08	**0.87**	**1.23**	-4.47	-3.06	-2.27	-3.46	-1.50	-1.03	2	2	4	
	Spline odds	0.43	5.24	7.60	**0.17**	4.00	5.92	0.34	7.23	7.68	**0.12**	1.48	1.94	**-0.15**	1.54	1.78	5	5	3	
	Spline normal	**-0.14**	3.67	6.97	-1.19	2.90	3.87	-0.97	5.68	6.04	-1.42	**-0.19**	**0.04**	-1.53	**0.19**	**0.49**	6	4	5	
	Selected parametric[3]	-7.48	6.67	10.46	-4.07	**0.36**	2.90	-3.70	10.22	4.31	-4.34	1.77	-1.68	-3.26	-1.52	-1.03	1	1	1	4
	Selected spline[4]	**-0.14**	**0.18**	6.97	-5.79	2.90	3.87	0.34	**0.87**	6.04	-1.42	**-0.19**	**0.04**	-1.53	**0.19**	**0.49**	8	7	7	11
Within 1 month (count of 9 models)		2	2	0	1	1	0	5	1	0	1	1	1	2	1	3				
Within CIs[1] (count of 9 models)		3	2	0	1	1	0	3	1	0	1	1	1	1	1	0				

Difference is calculated as predicted minus observed; [1] Observed value is within the predicted confidence intervals (CIs); [2] Comparison of the selected parametric and selected spline models only;

[3] Parametric model for each cohort selected based on lowest AIC of all parametric models; [4] Spline model for each cohort selected based on lowest AIC of all spline models

bold Prediction is closest to the observed value

green Gompertz plateaued early at a high percentage surviving and predictions appeared implausible

Table A2. Difference in **percentage surviving at 10 years** by cohort, model and follow-up duration (%)

		Cohort															Within 1% (count)	Within CIs[1] (count)	Closest to observed (count)	Closest selected model[2] (count)
		Breast			Colorectal			SCLC			NSCLC			Pancreatic						
		18-59	60-69	70+	18-59	60-69	70+	18-59	60-69	70+	18-59	60-69	70+	18-59	60-69	70+				
Observed percentage surviving (%)		14.67	8.50	4.51	6.51	4.01	2.05	2.09	0.91	0.32	2.56	1.30	0.50	1.40	0.43	0.21				
10 years of follow-up																				
Model	Exponential	-5.98	-4.57	-3.30	-5.67	-3.81	-2.04	-2.08	-0.91	-0.32	-2.54	-1.29	-0.50	-1.40	-0.43	-0.21	5	0	0	
	Weibull	-2.54	-0.96	-0.25	-3.96	-2.79	-1.58	-1.99	-0.88	-0.32	-2.32	-1.23	-0.48	-1.37	-0.43	-0.21	7	1	0	
	Gompertz	-0.06	0.27	0.73	-0.17	**-0.04**	0.33	-0.66	-0.18	**0.16**	-0.16	**0.10**	0.31	-0.02	**0.08**	0.21	15	8	4	
	Log-logistic	0.51	3.74	4.47	-1.23	0.29	0.95	-0.11	1.09	1.03	-0.80	0.15	0.57	-0.46	0.09	0.11	10	2	0	
	Log-normal	1.79	3.67	3.70	-1.21	-0.11	**0.17**	-0.53	0.33	0.26	-1.23	-0.40	0.02	-0.88	-0.26	-0.15	10	2	1	
	Generalized gamma	-1.30	-0.24	0.70	-2.17	-1.59	-0.55	-1.75	-0.74	-0.25	-1.54	-0.78	-0.26	-0.99	-0.39	-0.19	10	1	0	
	Spline hazard	**-0.02**	-0.05	**-0.03**	**-0.02**	-0.14	-0.23	-0.16	**0.04**	-0.25	**-0.04**	-0.11	-0.06	**0.00**	-0.09	**-0.05**	15	13	7	
	Spline odds	-0.02	0.04	0.33	-0.24	-0.46	-0.31	-0.03	-0.20	-0.24	-0.26	-0.21	**0.02**	-0.10	-0.18	-0.07	15	7	1	
	Spline normal	-0.02	**0.02**	0.13	-0.13	-0.31	-0.34	**0.01**	-0.07	-0.29	-0.14	-0.21	-0.08	-0.06	-0.17	-0.09	15	8	2	
	Selected parametric[3]	-0.06	-0.24	0.70	-0.17	**-0.04**	-0.55	-0.66	-0.18	-0.25	-0.80	0.15	-0.26	-0.46	0.09	0.11	15	4	1	2
	Selected spline[4]	**-0.02**	**0.02**	0.13	**-0.02**	-0.14	-0.31	**0.01**	**0.04**	-0.24	**-0.04**	-0.21	**0.02**	**0.00**	-0.09	-0.09	15	11	8	13
Within 1 % (count of 9 models)		5	6	6	4	6	7	6	8	8	5	7	9	7	9	9				
Within CIs[1] (count of 9 models)		5	5	4	4	3	0	4	2	0	3	2	3	4	2	1				
Censored at 20% surviving																				
Model	Exponential	-8.02	-5.97	-3.95	-6.35	-3.98	-2.05	-2.09	-0.91	-0.32	-2.56	-1.30	-0.50	-1.40	-0.43	-0.21	5	0	0	
	Weibull	-4.97	-2.06	-0.39	-6.25	-3.86	-1.99	-2.09	-0.91	-0.32	-2.56	-1.30	-0.50	-1.40	-0.43	-0.21	6	1	0	
	Gompertz	-0.71	2.99	9.46	-4.81	-1.11	8.87	-2.09	-0.91	8.12	-2.38	**-0.07**	4.15	-1.13	-0.39	1.06	4	2	1	
	Log-logistic	**0.28**	4.76	6.07	**-1.59**	0.58	1.43	0.28	1.78	1.72	**-0.85**	0.28	0.78	**-0.48**	0.19	0.16	9	1	4	
	Log-normal	1.61	5.02	5.70	-1.91	**0.09**	0.55	**-0.17**	0.97	0.81	-1.46	-0.38	**0.12**	-1.00	-0.23	-0.14	9	2	3	
	Generalized gamma	-3.62	-1.41	2.66	-5.71	-3.58	-0.86	-2.09	-0.91	**0.04**	-2.52	-1.25	-0.43	-1.40	-0.43	-0.21	6	1	1	
	Spline hazard	-1.43	-1.78	**-0.33**	-5.60	-3.79	-1.83	-2.09	-0.91	-0.32	-2.53	-1.29	-0.49	-1.40	-0.43	-0.21	6	1	1	
	Spline odds	-0.75	**0.22**	2.35	-3.27	-1.89	**0.08**	-1.92	**-0.79**	0.32	-1.41	-0.37	0.39	-0.77	**-0.12**	**0.06**	10	2	5	
	Spline normal	-0.96	-0.66	1.57	-4.39	-2.92	-0.96	-2.08	-0.91	-0.21	-2.14	-1.03	-0.24	-1.27	-0.40	-0.19	8	1	0	
	Selected parametric[3]	-0.71	-1.41	2.66	-5.71	-3.58	-0.86	-2.09	-0.91	**0.04**	-2.52	-1.25	-0.43	-1.40	-0.43	-0.21	7	2	1	3
	Selected spline[4]	-1.43	**0.22**	**-0.33**	-4.39	-2.92	-0.96	-1.92	-0.91	-0.32	-2.14	-1.03	-0.24	-1.27	-0.40	-0.19	8	2	2	11
Within 1 % (count of 9 models)		4	2	2	0	2	4	2	8	7	1	4	8	2	9	8				
Within CIs[1] (count of 9 models)		2	2	2	0	1	1	1	0	1	0	1	0	0	0	0				
Censored at 35% surviving																				
Model	Exponential	-10.13	-6.75	-4.26	-6.39	-3.99	-2.05	-2.09	-0.91	-0.32	-2.56	-1.30	-0.50	-1.40	-0.43	-0.21	5	0	0	
	Weibull	-8.03	-2.12	-0.36	-6.39	-3.89	-2.02	-2.09	-0.90	-0.32	-2.56	-1.30	-0.50	-1.40	-0.43	-0.21	6	1	0	
	Gompertz	-4.20	10.12	22.65	-6.28	1.90	18.16	-2.09	2.14	23.70	-2.56	-0.88	3.59	-1.40	-0.43	-0.21	3	0	0	
	Log-logistic	**0.77**	6.69	7.76	-0.97	1.55	1.89	1.53	3.19	2.40	**-0.63**	0.51	0.88	**-0.37**	0.31	0.19	8	0	3	
	Log-normal	3.04	7.74	7.96	**-0.80**	1.54	1.19	**1.48**	2.62	1.51	-1.09	**-0.05**	0.28	-0.86	-0.12	-0.12	6	1	3	
	Generalized gamma	-8.64	-0.43	6.70	-6.19	-3.43	**-0.11**	-2.09	-0.89	1.95	-2.55	-1.27	-0.46	-1.40	-0.43	-0.21	6	2	1	
	Spline hazard	-6.20	-4.02	**0.28**	-6.39	-3.89	-1.87	-2.09	-0.91	**-0.12**	-2.56	-1.29	-0.49	-1.40	-0.43	-0.21	6	2	2	
	Spline odds	-2.04	1.03	5.64	-3.42	**-0.91**	1.33	-1.81	**-0.32**	3.18	-1.22	0.15	0.91	-0.68	**0.12**	**0.09**	7	2	4	
	Spline normal	-3.15	**-0.21**	3.79	-4.77	-2.42	-0.21	-2.07	-0.84	1.64	-2.15	-0.80	**-0.11**	-1.29	-0.37	-0.19	7	2	2	
	Selected parametric[3]	-8.03	-0.43	6.70	-6.19	-3.43	**-0.11**	-2.09	-0.90	1.95	-2.55	-1.27	0.88	-1.40	-0.43	-0.21	6	2	1	2
	Selected spline[4]	-3.15	**-0.21**	**0.28**	-4.77	-2.42	-0.21	-2.07	-0.91	1.64	-2.15	-0.80	**-0.11**	-1.29	-0.37	-0.19	8	3	3	13
Within 1 % (count of 9 models)		1	2	2	2	1	2	0	6	3	1	5	8	3	9	9				
Within CIs[1] (count of 9 models)		0	2	2	0	0	2	0	0	1	0	2	0	0	1	0				
Censored at 50% surviving																				
Model	Exponential	-10.71	-7.19	-4.44	-6.39	-4.00	-2.05	-2.08	-0.91	-0.32	-2.56	-1.30	-0.50	-1.40	-0.43	-0.21	5	0	0	
	Weibull	-8.46	**-0.83**	**-1.10**	-6.40	-3.89	-2.04	-2.06	**-0.84**	-0.32	-2.56	-1.30	-0.50	-1.40	-0.43	-0.21	6	1	3	
	Gompertz	-4.93	27.41	37.59	-6.22	12.75	28.43	-2.09	30.00	30.57	-2.56	-1.30	-0.46	-1.40	-0.43	-0.21	3	1	0	
	Log-logistic	2.68	9.66	8.84	**0.16**	2.74	1.98	3.85	5.24	1.93	**-0.38**	0.61	0.88	**-0.19**	0.41	0.23	7	1	3	
	Log-normal	6.64	11.92	9.95	1.40	3.70	1.75	4.80	5.37	1.34	-0.42	**0.38**	0.49	-0.51	**0.07**	**-0.06**	6	1	3	
	Generalized gamma	-12.28	7.12	12.25	-5.88	-2.31	**0.93**	-2.09	4.14	1.84	-2.56	-1.29	-0.48	-1.40	-0.43	-0.21	4	0	1	
	Spline hazard	-7.85	-1.64	2.70	-6.38	-3.74	-1.79	-2.09	-0.88	**-0.23**	-2.56	-1.30	-0.50	-1.40	-0.43	-0.21	5	1	1	
	Spline odds	**-0.78**	6.18	9.30	-1.93	1.51	3.75	**-1.31**	2.98	3.87	-0.86	0.48	1.01	-0.53	0.34	0.50	6	1	2	
	Spline normal	-2.07	3.79	8.13	-3.64	**-0.05**	1.64	-1.97	1.33	2.22	-2.01	-0.72	**-0.14**	-1.25	-0.32	-0.11	5	1	2	
	Selected parametric[3]	-12.28	7.12	12.25	-5.88	-2.31	**0.93**	-2.09	4.14	1.34	-2.56	0.61	-0.48	-1.40	-0.43	-0.21	5	0	1	4
	Selected spline[4]	-2.07	-1.64	8.13	-6.38	**-0.05**	1.64	**-1.31**	-0.88	2.22	-2.01	-0.72	**-0.14**	-1.25	-0.32	-0.11	6	2	3	11
Within 1 % (count of 9 models)		1	1	0	1	1	1	0	3	3	3	4	8	3	9	9				
Within CIs[1] (count of 9 models)		1	2	0	1	1	0	0	0	0	0	0	1	0	1	0				

Difference is calculated as predicted minus observed; [1] Observed value is within the predicted confidence intervals (CIs); [2] Comparison of the selected parametric and selected spline models only;
[3] Parametric model for each cohort selected based on lowest AIC of all parametric models; [4] Spline model for each cohort selected based on lowest AIC of all spline models

bold Prediction is closest to the observed value

green Gompertz plateaued early at a high percentage surviving and predictions appeared implausible

Table B1. Number of internal knots selected for each spline model

		Cohort															Totals		
Follow up duration	**Model scale**	**Breast**			**Colorectal**			**SCLC**			**NSCLC**			**Pancreatic**					
		18-59	**60-69**	**70+**	**18-59**	**60-69**	**70+**	**18-59**	**60-69**	**70+**	**18-59**	**60-69**	**70+**	**18-59**	**60-69**	**70+**	**1 knot**	**2 knots**	**3 knots**
10 years of follow-up	Hazard	3*	3	3	3*	3*	3	2	3*	3	3*	3	3	3*	3*	3	0	1	14
	Odds	3	3	3	3	3	3*	3	3	2*	3	3*	2*	3	3	2	0	3	12
	Normal	3	3*	3*	3	3	3	3*	3	2	3	3	3	3	3	3*	0	1	14
Censored at 20% surviving	Hazard	3*	3	3*	3	3	2	3	2*	3*	3	3	3	3	2	3	0	3	12
	Odds	3	3*	3	3	3	3	3*	3	3	2	3	3	3	3	2	0	2	13
	Normal	3	3	2	3*	2*	3*	3	3	3	3*	3*	3*	3*	1*	3*	1	2	12
Censored at 35% surviving	Hazard	3	3	2*	2	2	3	3	3*	3	3	3	2	3	3	2	0	5	10
	Odds	3	2	2	3	3	3	2	3	3	2	3	1	3	3	3	1	4	10
	Normal	3*	2*	3	2*	3*	3*	2*	3	3*	1*	2*	2*	1*	3*	1*	3	5	7
Censored at 50% surviving	Hazard	3	2*	2	2*	2	1	3	3*	3	2	2	3	3	3	3	1	6	8
	Odds	3	2	3	3	3	1	3*	3	3	2	1	3	3	3	3	2	2	11
	Normal	3*	3	3*	3	2*	3*	3	3	3*	2*	1*	1*	3*	3*	3*	2	2	11
Totals																			
All follow-up durations	1 knot	0	0	0	0	0	2	0	0	0	1	2	2	1	1	1	10		
	2 knots	0	4	4	3	4	1	3	1	2	5	2	3	0	1	3		36	
	3 knots	12	8	8	9	8	9	9	11	10	6	8	7	11	10	8			134
Right censored durations only	1 knot	0	0	0	0	0	2	0	0	0	1	2	2	1	1	1	10		
	2 knots	0	4	4	3	4	1	2	1	0	5	2	2	0	1	2		31	
	3 knots	9	5	5	6	5	6	7	8	9	3	5	5	8	7	6			94

* Indicates the spline model with lowest AIC overall for the cohort at that follow-up duration

Table B2. Models selected based on lowest AIC

		Cohort															Model (count)								
Follow up duration	**Model type**	**Breast**			**Colorectal**			**SCLC**			**NSCLC**			**Pancreatic**			Exponential	Weibull	Gompertz	Log-logistic	Log-normal	Generalized gamma	Spline hazard	Spline odds	Spline normal
		18-59	**60-69**	**70+**	**18-59**	**60-69**	**70+**	**18-59**	**60-69**	**70+**	**18-59**	**60-69**	**70+**	**18-59**	**60-69**	**70+**									
10 years of follow-up	Selected parametric	gompertz	generalized gamma	generalized gamma	gompertz	gompertz	generalized gamma	gompertz	gompertz	generalized gamma	log-logistic	log-logistic	generalized gamma	log-logistic	log-logistic	log-logistic	0	0	5	5	0	5			
	Selected spline	hazard, 3 knots	normal, 3 knots	normal, 3 knots	hazard, 3 knots	hazard, 3 knots	odds, 3 knots	normal, 3 knots	hazard, 3 knots	odds, 2 knots	hazard, 3 knots	odds, 3 knots	odds, 2 knots	hazard, 3 knots	hazard, 3 knots	normal, 3 knots							7	4	4
Censored at 20% surviving	Selected parametric	gompertz	generalized gamma	generalized gamma	generalized gamma	generalized gamma	generalized gamma	generalized gamma	generalized gamma	generalized gamma	generalized gamma	generalized gamma	generalized gamma	generalized gamma	generalized gamma	generalized gamma	0	0	1	0	0	14			
	Selected spline	hazard, 3 knots	odds, 3 knots	hazard, 3 knots	normal, 3 knots	normal, 2 knots	normal, 3 knots	odds, 3 knots	hazard, 2 knots	hazard, 3 knots	normal, 3 knots	normal, 3 knots	normal, 3 knots	normal, 3 knots	normal, 1 knots	normal, 3 knots							4	2	9
Censored at 35% surviving	Selected parametric	weibull	generalized gamma	generalized gamma	generalized gamma	generalized gamma	generalized gamma	generalized gamma	weibull	generalized gamma	generalized gamma	generalized gamma	log-logistic	generalized gamma	generalized gamma	generalized gamma	0	2	0	1	0	12			
	Selected spline	normal, 3 knots	normal, 2 knots	hazard, 2 knots	normal, 2 knots	normal, 3 knots	normal, 3 knots	normal, 2 knots	hazard, 3 knots	normal, 3 knots	normal, 1 knots	normal, 2 knots	normal, 2 knots	normal, 1 knots	normal, 3 knots	normal, 1 knots							2	0	13
Censored at 50% surviving	Selected parametric	generalized gamma	generalized gamma	generalized gamma	generalized gamma	generalized gamma	generalized gamma	generalized gamma	generalized gamma	log-normal	generalized gamma	log-logistic	generalized gamma	generalized gamma	generalized gamma	generalized gamma	0	0	0	1	1	13			
	Selected spline	normal, 3 knots	hazard, 2 knots	normal, 3 knots	hazard, 2 knots	normal, 2 knots	normal, 3 knots	odds, 3 knots	hazard, 3 knots	normal, 3 knots	normal, 2 knots	normal, 1 knots	normal, 1 knots	normal, 3 knots	normal, 3 knots	normal, 3 knots							3	1	11

Figure C1. Model predictions of restricted mean survival time (RMST) across the 15 cohorts for (a) the observed portion of the data and (b) the extrapolated (unobserved) tail of the data

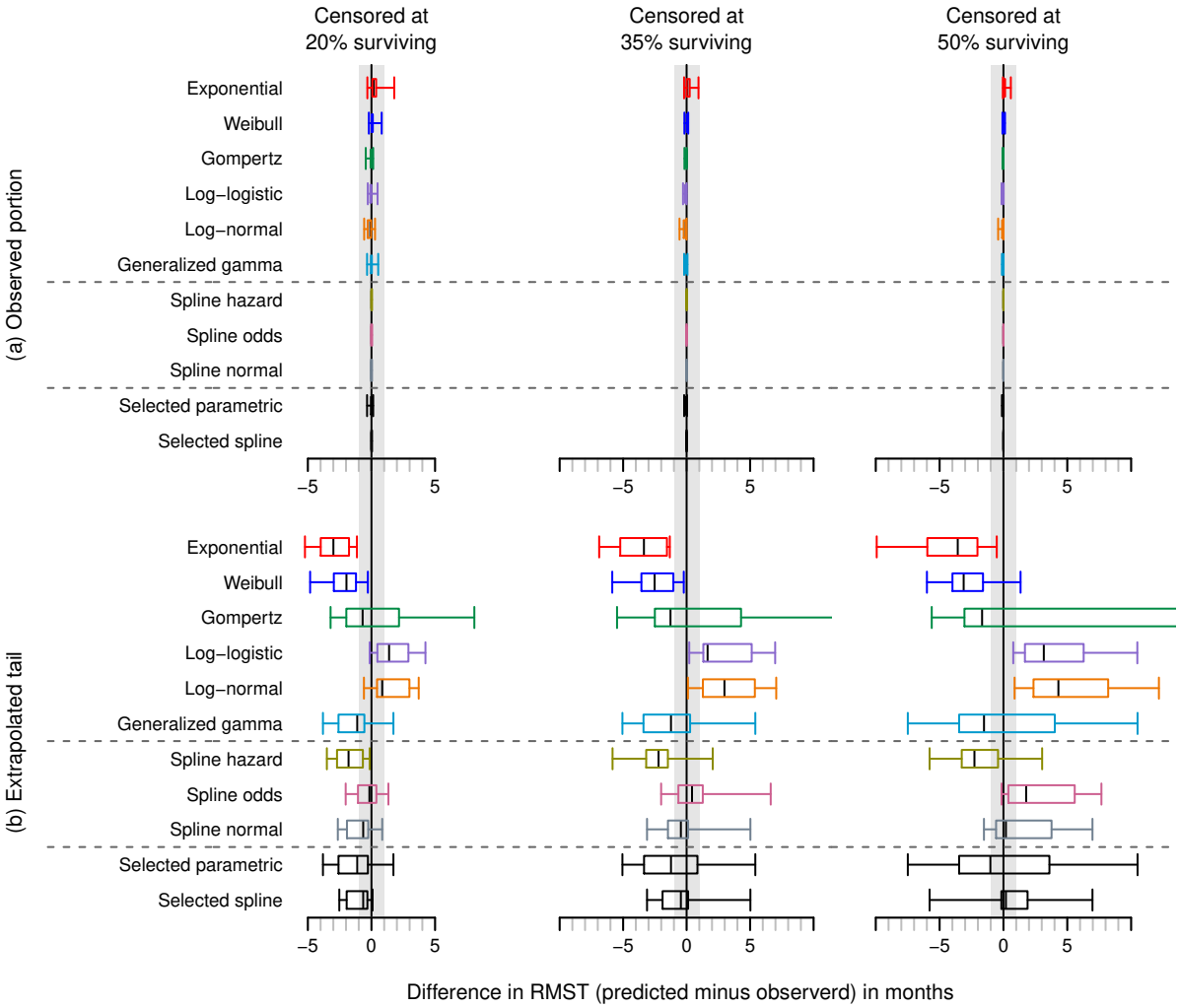


Table C1. Difference in **restricted mean survival time (RMST)** for the **observed portion** of the data by cohort, model and follow-up duration (months)

		Cohort															Within 1 month (count)	Within CIs[1] (count)	Closest to observed (count)	Closest selected model[2]
		Breast			Colorectal			SCLC			NSCLC			Pancreatic						
		18-59	60-69	70+	18-59	60-69	70+	18-59	60-69	70+	18-59	60-69	70+	18-59	60-69	70+				
Observed RMST at 10 years (months)		42.00	33.96	26.00	23.53	18.59	13.10	12.87	10.38	6.78	13.08	10.55	7.88	8.67	5.65	3.89				
Censored at 20% surviving																				
Observed RMST at 20% surviving (months)		35.44	26.13	18.52	14.90	11.67	7.58	8.53	6.95	4.67	7.44	6.17	4.62	4.48	3.36	2.19				
Model	Exponential	1.79	1.21	1.25	0.29	0.30	0.44	-0.32	-0.08	0.20	0.12	0.15	0.17	0.09	0.04	0.06	12	1	0	
	Weibull	0.80	-0.01	0.02	0.18	0.06	0.12	-0.21	-0.17	0.01	0.12	0.09	0.09	0.08	0.04	0.04	15	5	0	
	Gompertz	0.14	-0.25	-0.45	0.03	-0.03	-0.08	-0.01	-0.05	-0.08	0.03	0.01	0.00	0.02	0.01	0.01	15	13	0	
	Log-logistic	0.48	-0.08	-0.29	0.03	-0.04	-0.07	-0.10	-0.17	-0.11	0.01	-0.01	-0.01	0.01	-0.01	0.00	15	13	1	
	Log-normal	0.28	-0.57	-0.48	-0.15	-0.21	-0.09	-0.45	-0.38	-0.13	-0.04	-0.05	-0.02	-0.02	-0.04	-0.01	15	8	0	
	Generalized gamma	0.53	-0.12	-0.35	0.05	-0.02	-0.05	-0.02	0.08	-0.10	0.04	0.01	0.01	0.02	0.00	0.01	15	13	0	
	Spline hazard	0.04	0.02	0.00	0.01	0.01	0.01	0.00	0.01	-0.01	0.01	0.00	0.00	0.00	0.00	0.00	15	15	5	
	Spline odds	0.05	0.01	-0.03	0.01	0.00	0.00	0.00	0.01	-0.02	0.00	0.00	0.00	0.00	0.00	0.00	15	15	6	
	Spline normal	0.04	0.02	-0.04	0.01	0.01	0.00	0.01	0.01	-0.02	0.00	0.00	0.00	0.00	-0.01	0.00	15	15	3	
	Selected parametric[3]	0.14	-0.12	-0.35	0.05	-0.02	-0.05	-0.02	0.08	-0.10	0.04	0.01	0.01	0.02	0.00	0.01	15	13	0	1
	Selected spline[4]	0.04	0.01	0.00	0.01	0.01	0.00	0.00	0.01	-0.01	0.00	0.00	0.00	0.00	-0.01	0.00	15	15	7	14
Within 1 month (count of 9 models)		8	8	8	9	9	9	9	9	9	9	9	9	9	9	9				
Within CIs[1] (count of 9 models)		7	7	5	7	7	5	6	6	5	7	7	7	7	8	7				
Censored at 35% surviving																				
Observed RMST at 35% surviving (months)		25.23	19.31	13.07	11.52	9.04	5.40	7.38	5.92	3.58	5.80	4.74	3.43	3.46	2.64	1.67				
Model	Exponential	0.43	0.76	0.94	0.02	0.18	0.27	-0.20	0.06	0.22	-0.01	0.05	0.07	0.00	0.01	0.01	15	6	1	
	Weibull	0.03	-0.09	0.12	0.02	0.02	0.08	-0.17	-0.07	0.08	0.04	0.05	0.05	0.02	0.02	0.02	15	9	0	
	Gompertz	0.01	-0.15	-0.17	0.00	-0.02	-0.02	-0.02	-0.04	-0.02	0.01	0.01	0.01	0.01	0.00	0.00	15	15	0	
	Log-logistic	-0.11	-0.28	-0.14	-0.08	-0.08	-0.02	-0.18	-0.13	-0.01	-0.01	-0.01	0.00	-0.01	-0.01	0.00	15	13	0	
	Log-normal	-0.52	-0.56	-0.21	-0.24	-0.19	-0.04	-0.37	-0.22	0.00	-0.08	-0.05	-0.02	-0.05	-0.03	-0.02	15	1	1	
	Generalized gamma	0.07	-0.19	-0.17	-0.02	-0.04	-0.02	-0.06	-0.08	-0.01	0.01	0.01	0.01	0.00	0.00	0.00	15	15	0	
	Spline hazard	0.00	0.01	0.01	0.00	0.00	0.00	0.00	0.00	-0.01	0.00	0.00	0.00	0.00	0.00	0.00	15	15	7	
	Spline odds	0.00	0.00	-0.01	0.00	0.00	0.00	0.00	0.00	-0.01	0.00	0.00	0.00	0.00	0.00	0.00	15	15	4	
	Spline normal	-0.01	0.00	0.00	-0.01	-0.01	0.00	0.00	-0.01	-0.01	0.00	-0.01	0.00	0.00	0.00	0.00	15	15	2	
	Selected parametric[3]	0.03	-0.19	-0.17	-0.02	-0.04	-0.02	-0.06	-0.07	-0.01	0.01	0.01	0.00	0.00	0.00	0.00	15	15	0	1
	Selected spline[4]	-0.01	0.00	0.01	-0.01	-0.01	0.00	0.00	0.00	-0.01	0.00	-0.01	0.00	0.00	0.00	0.00	15	15	2	14
Within 1 month (count of 9 models)		9	9	9	9	9	9	9	9	9	9	9	9	9	9	9				
Within CIs[1] (count of 9 models)		7	7	7	8	7	6	5	7	7	8	6	6	8	8	7				
Censored at 50% surviving																				
Observed RMST at 50% surviving (months)		18.58	13.85	8.31	8.83	6.68	3.61	6.18	4.64	2.23	4.41	3.50	2.46	2.64	1.98	1.25				
Model	Exponential	0.13	0.58	0.51	0.01	0.13	0.13	-0.02	0.15	0.08	-0.04	0.00	0.01	-0.02	-0.01	-0.01	15	8	1	
	Weibull	-0.08	0.04	0.12	0.02	0.03	0.05	-0.08	0.03	0.04	0.01	0.02	0.02	0.01	0.01	0.01	15	10	0	
	Gompertz	-0.01	-0.07	-0.05	0.00	-0.01	0.00	-0.02	-0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	15	15	2	
	Log-logistic	-0.15	-0.09	0.01	-0.04	-0.02	0.01	-0.10	-0.02	0.02	-0.01	0.00	0.00	-0.01	-0.01	0.00	15	14	0	
	Log-normal	-0.41	-0.21	-0.03	-0.13	-0.08	0.00	-0.19	-0.05	0.01	-0.06	-0.03	-0.02	-0.04	-0.03	-0.01	15	4	0	
	Generalized gamma	0.01	-0.12	-0.05	-0.01	-0.02	0.00	-0.02	-0.04	0.01	0.00	0.00	0.00	0.00	0.00	0.00	15	15	1	
	Spline hazard	0.00	0.00	-0.01	0.00	0.00	0.00	0.00	-0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	15	15	7	
	Spline odds	0.00	-0.01	0.00	0.00	0.00	0.00	0.00	-0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	15	15	2	
	Spline normal	-0.01	0.00	-0.01	-0.01	0.00	0.00	0.00	-0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	15	15	2	
	Selected parametric[3]	0.01	-0.12	-0.05	-0.01	-0.02	0.00	-0.02	-0.04	0.01	0.00	0.00	0.00	0.00	0.00	0.00	15	15	1	1
	Selected spline[4]	-0.01	0.00	-0.01	0.00	0.00	0.00	0.00	-0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	15	15	5	14
Within 1 month (count of 9 models)		9	9	9	9	9	9	9	9	9	9	9	9	9	9	9				
Within CIs[1] (count of 9 models)		8	7	7	8	7	7	7	8	7	7	7	7	8	8	8				

Difference is calculated as predicted minus observed; [1] Observed value is within the predicted confidence intervals (CIs); [2] Comparison of the selected parametric and selected spline models only;
[3] Parametric model for each cohort selected based on lowest AIC of all parametric models; [4] Spline model for each cohort selected based on lowest AIC of all spline models

bold Prediction is closest to the observed value

green Gompertz plateaued early at a high percentage surviving and predictions appeared implausible

Table C2. Difference in **restricted mean survival time (RMST)** for the **extrapolated (unobserved) tail** of the data by cohort, model and follow-up duration (months)

		Cohort															Within 1 month (count)	Within CIs[1] (count)	Closest to observed (count)	Closest selected model[2]
		Breast			Colorectal			SCLC			NSCLC			Pancreatic						
		18-59	60-69	70+	18-59	60-69	70+	18-59	60-69	70+	18-59	60-69	70+	18-59	60-69	70+				
Observed RMST at 10 years (months)		42.00	33.96	26.00	23.53	18.59	13.10	12.87	10.38	6.78	13.08	10.55	7.88	8.67	5.65	3.89				
Censored at 20% surviving																				
Observed RMST at 20% surviving (months)		6.57	7.83	7.48	8.63	6.93	5.53	4.34	3.44	2.11	5.64	4.38	3.26	4.19	2.29	1.70				
Model	Exponential	-2.48	-3.36	-4.23	-5.23	-4.32	-4.08	-1.90	-1.62	-1.14	-3.90	-2.99	-2.29	-3.16	-1.49	-1.21	0	0	1	
	Weibull	-1.33	-0.69	-0.28	-4.82	-3.34	-2.64	-2.31	-1.27	-0.35	-3.90	-2.79	-1.97	-3.12	-1.49	-1.17	3	1	2	
	Gompertz	-0.19	1.08	4.31	-3.16	-0.69	6.71	-3.21	-1.78	8.09	-2.94	-0.71	3.25	-2.17	-1.07	0.65	4	1	1	
	Log-logistic	0.29	2.70	3.76	0.14	2.08	2.22	3.12	4.24	3.89	-0.03	0.96	1.38	-0.14	0.99	0.63	7	3	3	
	Log-normal	0.85	2.94	3.71	0.40	2.13	1.73	3.16	3.68	3.03	-0.25	0.56	0.81	-0.59	0.48	0.10	8	0	1	
	Generalized gamma	-0.91	-0.37	1.73	-3.81	-2.61	-0.15	-3.46	-2.57	1.49	-3.12	-1.90	-0.88	-2.58	-1.12	-0.74	5	1	1	
	Spline hazard	-0.29	-0.58	-0.14	-3.51	-3.04	-1.80	-3.02	-2.29	-0.51	-2.94	-2.15	-1.21	-2.47	-1.19	-0.77	5	2	1	
	Spline odds	-0.16	0.09	1.15	-1.77	-1.03	0.46	-2.03	-1.35	1.33	-1.09	-0.27	0.62	-0.86	0.08	0.31	8	2	4	
	Spline normal	-0.19	-0.21	0.85	-2.53	-1.93	-0.56	-2.64	-1.93	0.35	-1.97	-1.22	-0.29	-1.74	-0.65	-0.34	8	2	1	
	Selected parametric[3]	-0.19	-0.37	1.73	-3.81	-2.61	-0.15	-3.46	-2.57	1.49	-3.12	-1.90	-0.88	-2.58	-1.12	-0.74	5	2	1	2
	Selected spline[4]	-0.29	0.09	-0.14	-2.53	-1.93	-0.56	-2.03	-2.29	-0.51	-1.97	-1.22	-0.29	-1.74	-0.65	-0.34	8	3	3	13
Within 1 month (count of 9 models)		7	5	3	2	1	3	0	0	3	2	4	4	3	4	7				
Within CIs[1] (count of 9 models)		5	3	2	1	0	0	0	0	0	1	0	0	0	0	0				
Censored at 35% surviving																				
Observed RMST at 35% surviving (months)		16.77	14.65	12.93	12.01	9.55	7.70	5.50	4.46	3.20	7.28	5.81	4.45	5.21	3.01	2.23				
Model	Exponential	-5.39	-5.56	-6.89	-5.86	-4.88	-5.07	-1.32	-1.35	-1.49	-4.16	-3.32	-2.68	-3.36	-1.61	-1.35	0	0	1	
	Weibull	-3.66	-0.67	-0.35	-5.86	-3.58	-3.44	-1.51	-0.23	-0.23	-4.52	-3.31	-2.52	-3.54	-1.72	-1.42	4	0	2	
	Gompertz	-1.89	5.37	14.18	-5.48	2.00	15.75	-3.11	3.20	24.56	-4.43	-1.79	2.66	-3.47	-1.55	-1.27	0	0	0	
	Log-logistic	1.39	5.01	5.72	1.25	3.74	3.03	5.77	6.98	5.24	0.52	1.51	1.65	0.21	1.41	0.74	3	0	2	
	Log-normal	3.13	6.01	6.31	2.31	4.58	2.98	6.65	7.06	4.74	0.85	1.57	1.40	0.10	1.16	0.34	3	0	2	
	Generalized gamma	-4.04	0.51	5.34	-5.06	-2.13	1.20	-4.04	0.04	5.41	-3.74	-2.10	-1.13	-3.05	-1.23	-0.98	3	1	1	
	Spline hazard	-2.51	-2.21	0.46	-5.83	-3.55	-2.01	-3.33	-2.04	2.06	-3.83	-2.52	-1.65	-3.04	-1.30	-1.13	1	0	0	
	Spline odds	-0.71	0.64	4.00	-2.01	0.41	2.31	-1.52	0.43	6.63	-0.70	0.79	1.72	-0.60	0.82	0.40	9	0	2	
	Spline normal	-1.15	0.09	2.82	-3.12	-1.03	0.77	-2.37	-0.71	5.02	-1.99	-0.45	0.15	-1.80	-0.31	-0.42	7	1	5	
	Selected parametric[3]	-3.66	0.51	5.34	-5.06	-2.13	1.20	-4.04	-0.23	5.41	-3.74	-2.10	1.65	-3.05	-1.23	-0.98	3	0	0	1
	Selected spline[4]	-1.15	0.09	0.46	-3.12	-1.03	0.77	-2.37	-2.04	5.02	-1.99	-0.45	0.15	-1.80	-0.31	-0.42	7	1	5	14
Within 1 month (count of 9 models)		1	4	2	0	1	1	0	4	1	3	2	1	3	2	5				
Within CIs[1] (count of 9 models)		0	1	0	0	0	0	0	1	0	0	0	0	0	0	0				
Censored at 50% surviving																				
Observed RMST at 50% surviving (months)		23.42	20.11	17.69	14.70	11.91	9.49	6.70	5.74	4.54	8.67	7.06	5.42	6.03	3.67	2.64				
Model	Exponential	-6.44	-7.22	-9.93	-5.91	-5.37	-6.01	-0.52	-1.26	-2.40	-4.22	-3.58	-2.98	-3.39	-1.67	-1.40	1	0	0	
	Weibull	-4.07	0.84	-1.60	-5.99	-3.62	-4.59	0.39	1.34	-1.67	-5.00	-3.92	-3.12	-3.79	-1.98	-1.60	2	0	1	
	Gompertz	-2.29	17.92	28.10	-5.26	11.35	26.51	-0.71	30.53	32.13	-5.62	-3.75	-2.36	-4.09	-2.22	-1.67	1	0	0	
	Log-logistic	3.43	8.11	6.93	3.05	5.63	3.16	9.93	10.49	4.16	1.15	1.77	1.63	0.77	1.73	0.93	2	0	0	
	Log-normal	6.76	10.26	8.58	5.57	7.83	4.05	12.18	11.69	4.30	2.59	2.77	2.12	1.42	2.08	0.87	1	0	0	
	Generalized gamma	-7.49	6.79	10.50	-4.05	0.38	2.90	-3.69	10.26	5.14	-4.34	-2.63	-1.68	-3.26	-1.52	-1.02	1	0	1	
	Spline hazard	-3.59	0.18	3.03	-5.78	-2.58	-1.55	-3.08	0.88	1.23	-4.47	-3.06	-2.27	-3.46	-1.50	-1.02	2	1	4	
	Spline odds	0.44	5.25	7.60	0.17	4.00	5.92	0.33	7.24	7.67	0.12	1.48	1.94	-0.15	1.54	1.78	5	0	4	
	Spline normal	-0.14	3.66	6.97	-1.18	2.90	3.87	-0.97	5.68	6.04	-1.42	-0.19	0.04	-1.53	0.19	0.49	6	1	5	
	Selected parametric[3]	-7.49	6.79	10.50	-4.05	0.38	2.90	-3.69	10.26	4.30	-4.34	1.77	-1.68	-3.26	-1.52	-1.02	1	0	1	4
	Selected spline[4]	-0.14	0.18	6.97	-5.78	2.90	3.87	0.33	0.88	6.04	-1.42	-0.19	0.04	-1.53	0.19	0.49	8	2	8	11
Within 1 month (count of 9 models)		2	2	0	1	1	0	5	1	0	1	1	1	2	1	3				
Within CIs[1] (count of 9 models)		1	1	0	0	0	0	0	0	0	0	0	0	0	0	0				

Difference is calculated as predicted minus observed; [1] Observed value is within the predicted confidence intervals (CIs); [2] Comparison of the selected parametric and selected spline models only;
[3] Parametric model for each cohort selected based on lowest AIC of all parametric models; [4] Spline model for each cohort selected based on lowest AIC of all spline models

bold Prediction is closest to the observed value

green Gompertz plateaued early at a high percentage surviving and predictions appeared implausible

Figure D1. Breast 18–59 years cohort (All models)

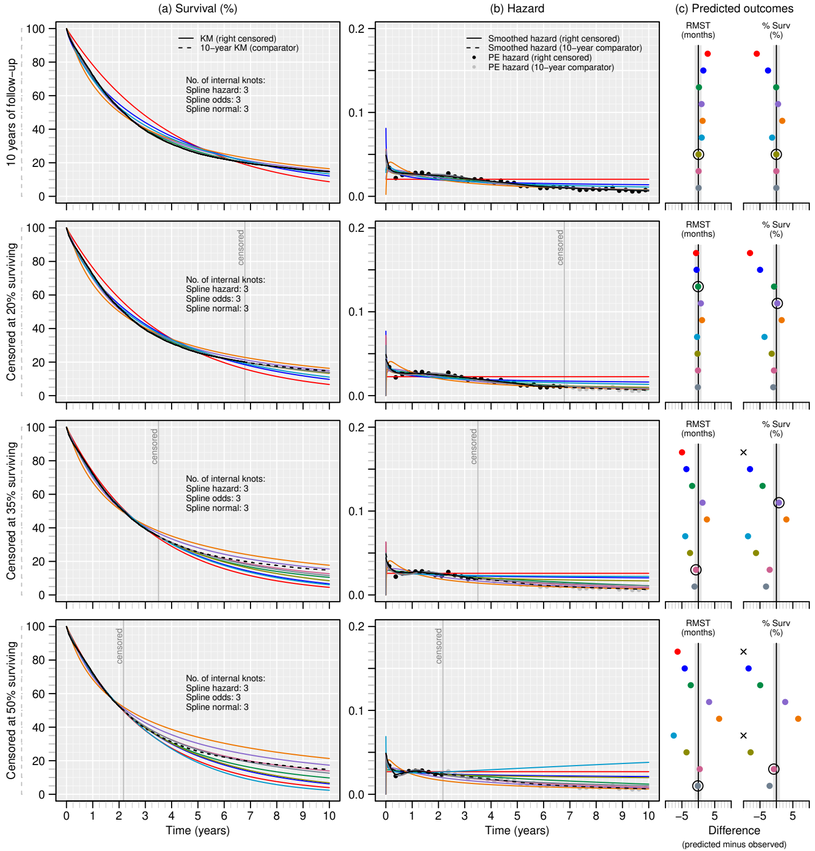


Figure D2. Breast 60–69 years cohort (All models)

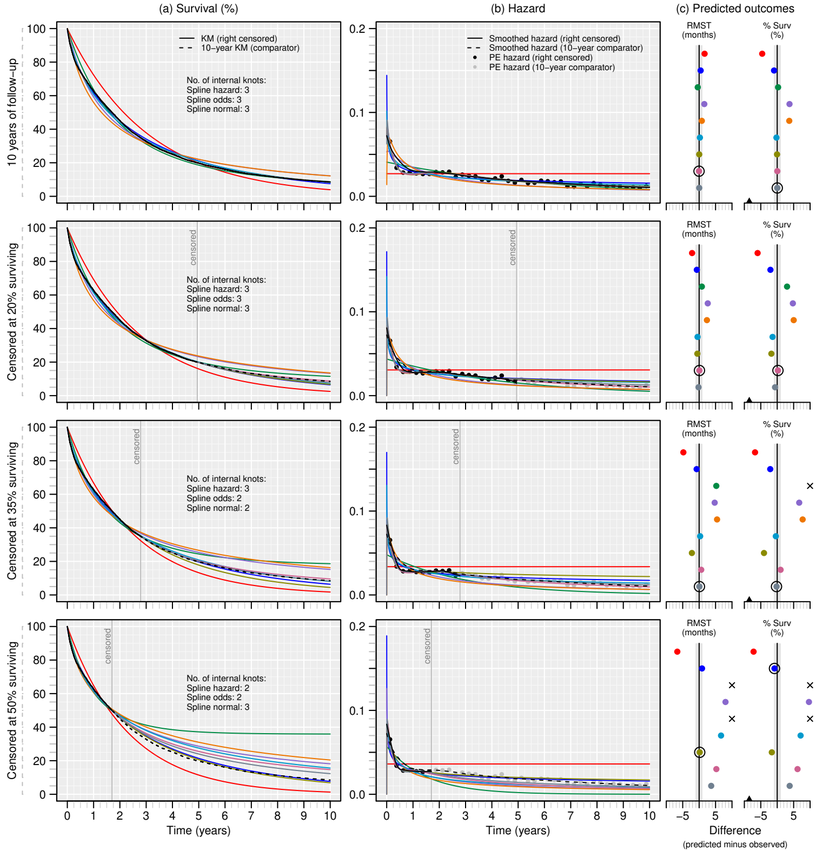


Figure D3. Breast 70+ years cohort (All models)

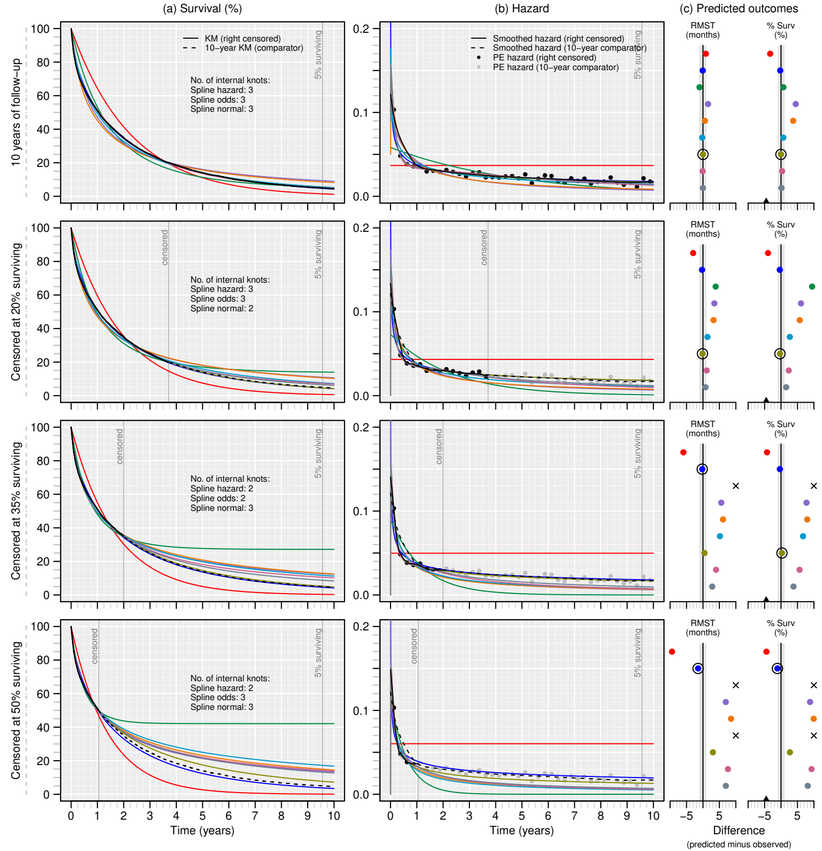


Figure E1. Colorectal 18–59 years cohort
(All models)

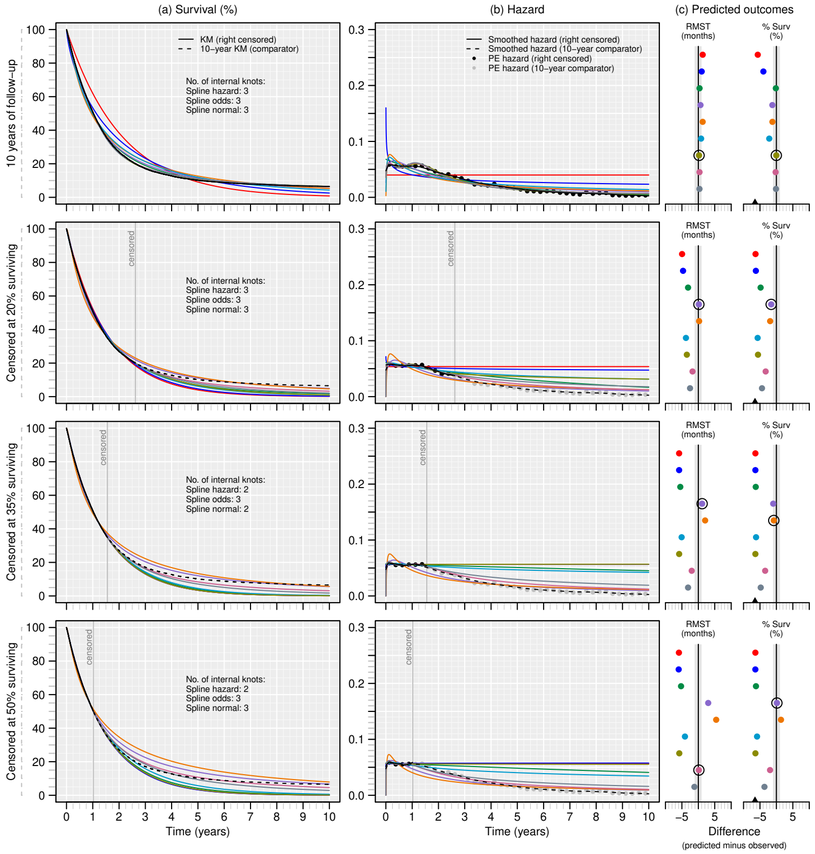


Figure E2. Colorectal 60–69 years cohort
(All models)

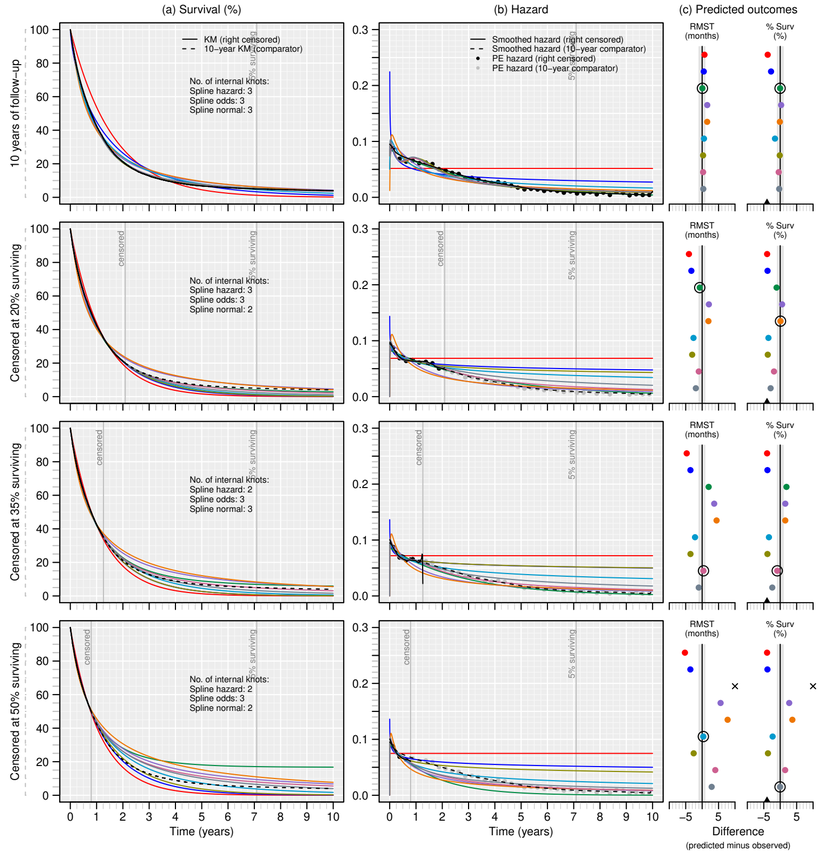


Figure E3. Colorectal 70+ years cohort
(All models)

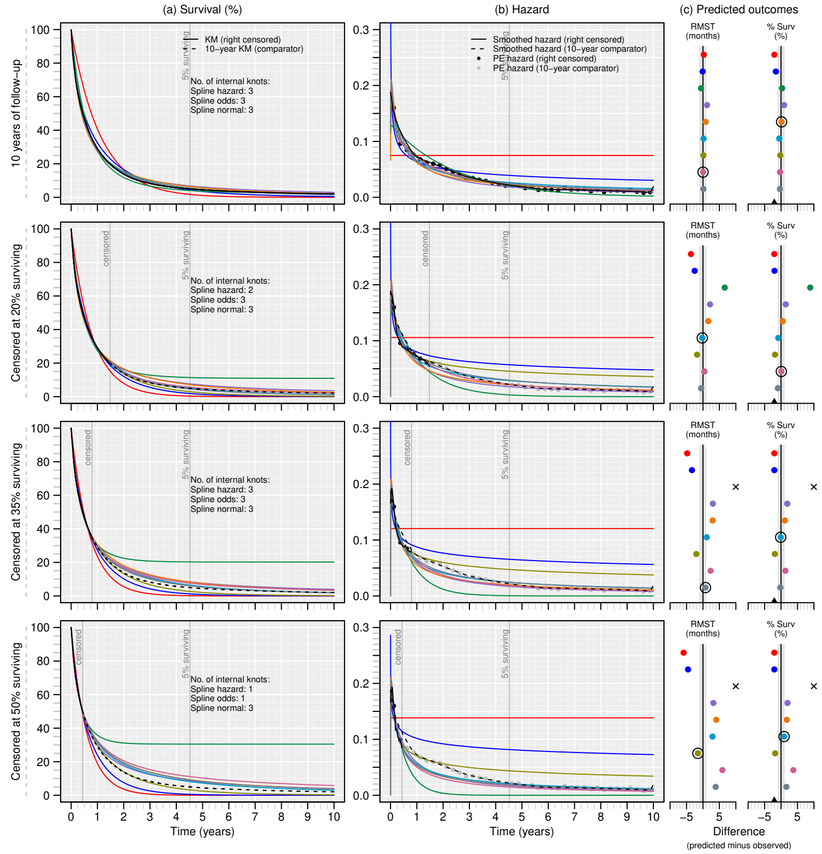
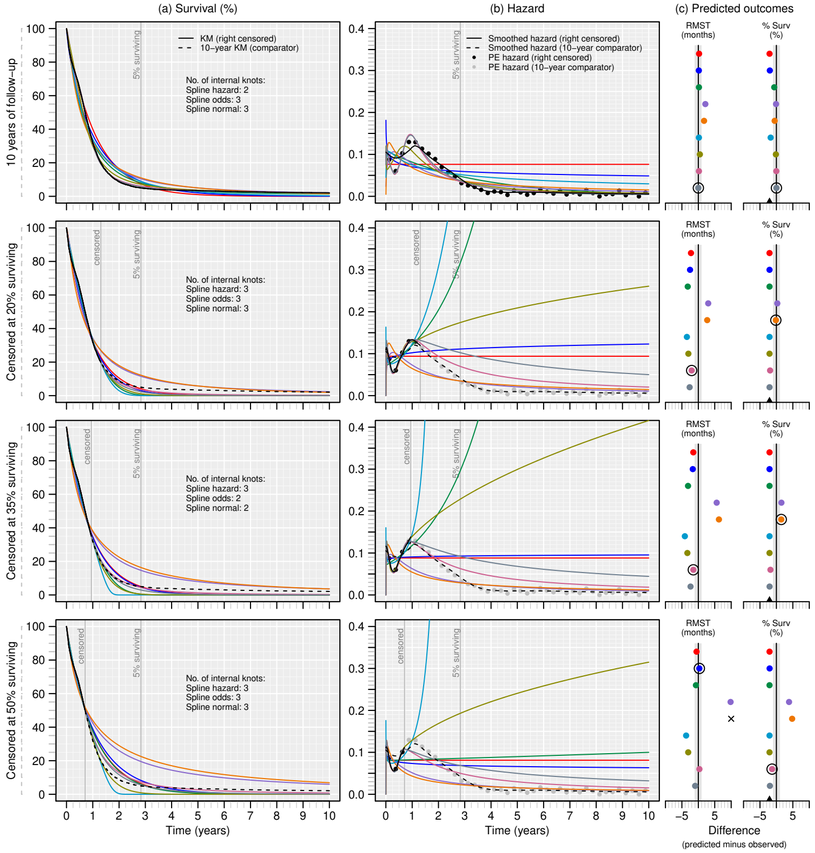
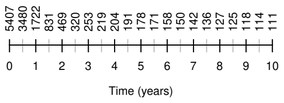


Figure F1. SCLC 18–59 years cohort (All models)



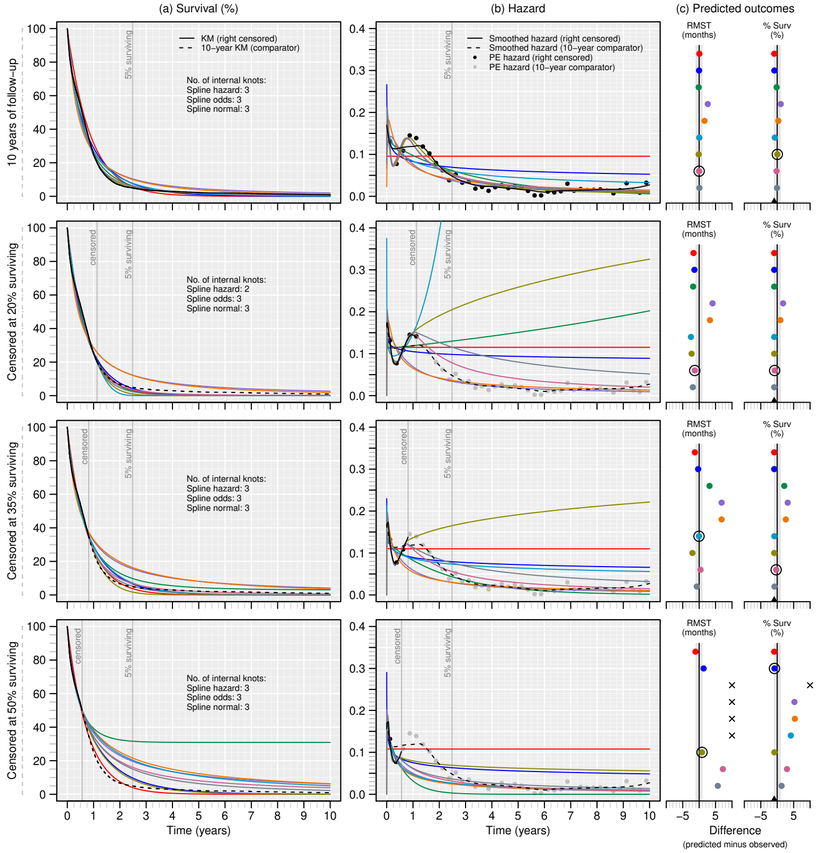
(d) Number at risk



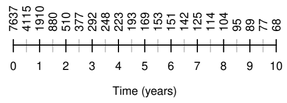
(e) 10–year observed outcomes

RMST = 12.87 months; percentage surviving = 2.09%

Figure F2. SCLC 60–69 years cohort (All models)



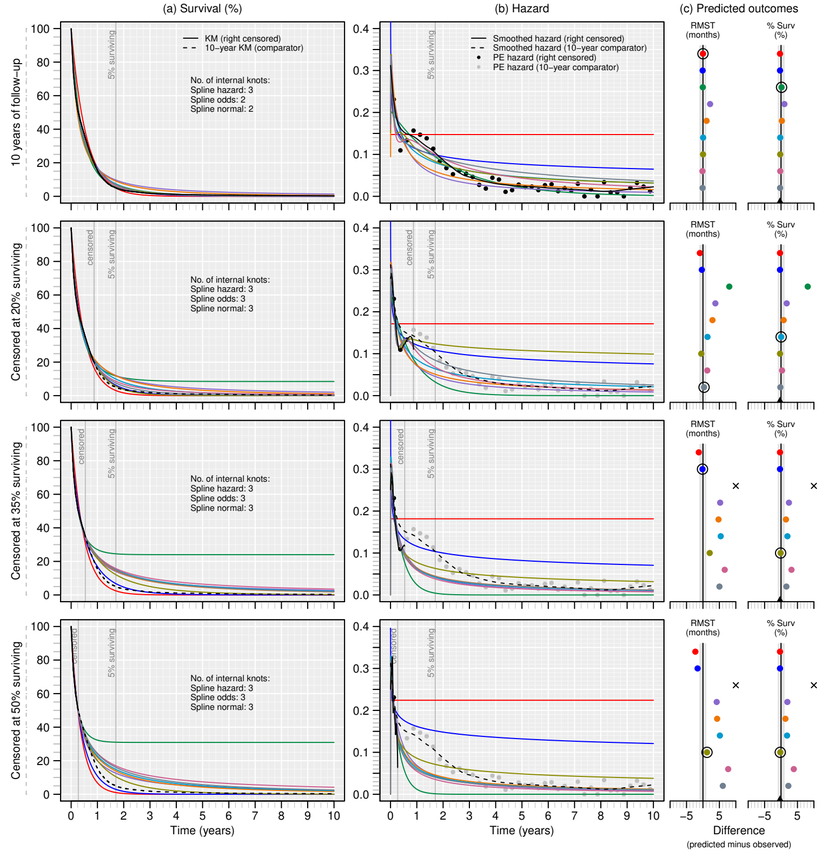
(d) Number at risk



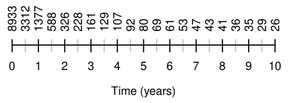
(e) 10–year observed outcomes

RMST = 10.38 months; percentage surviving = 0.91%

Figure F3. SCLC 70+ years cohort (All models)



(d) Number at risk



(e) 10–year observed outcomes

RMST = 6.78 months; percentage surviving = 0.32%

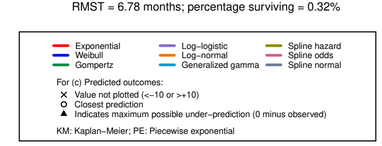
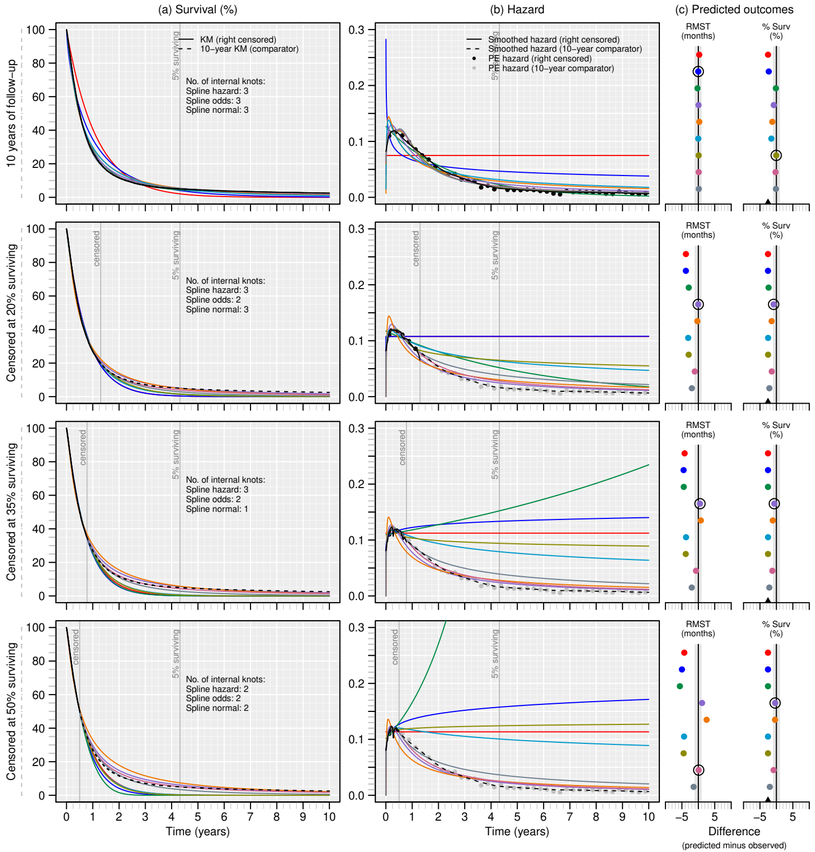
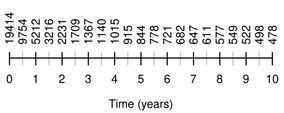


Figure G1. NSCLC 18–59 years cohort
(All models)



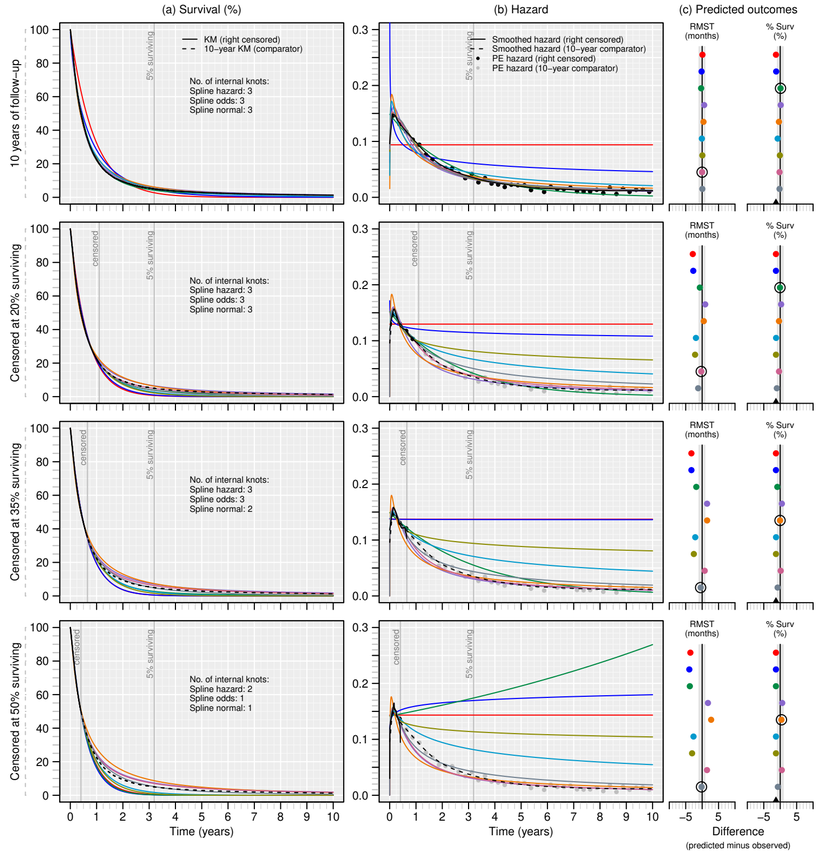
(d) Number at risk



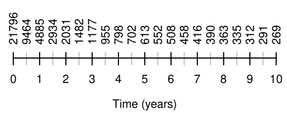
(e) 10−year observed outcomes

RMST = 13.08 months; percentage surviving = 2.56%

Figure G2. NSCLC 60–69 years cohort
(All models)



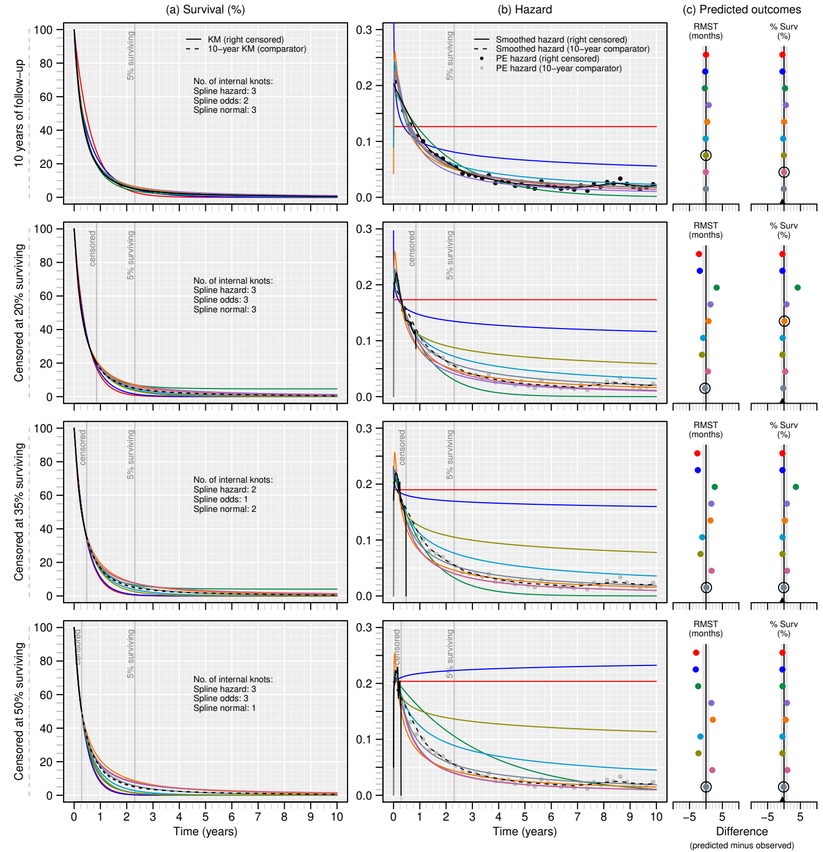
(d) Number at risk



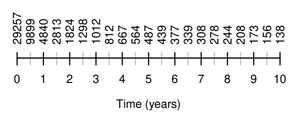
(e) 10−year observed outcomes

RMST = 10.55 months; percentage surviving = 1.30%

Figure G3. NSCLC 70+ years cohort
(All models)



(d) Number at risk



(e) 10−year observed outcomes

RMST = 7.88 months; percentage surviving = 0.50%

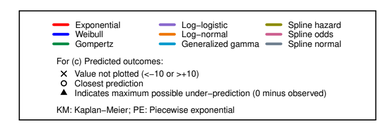
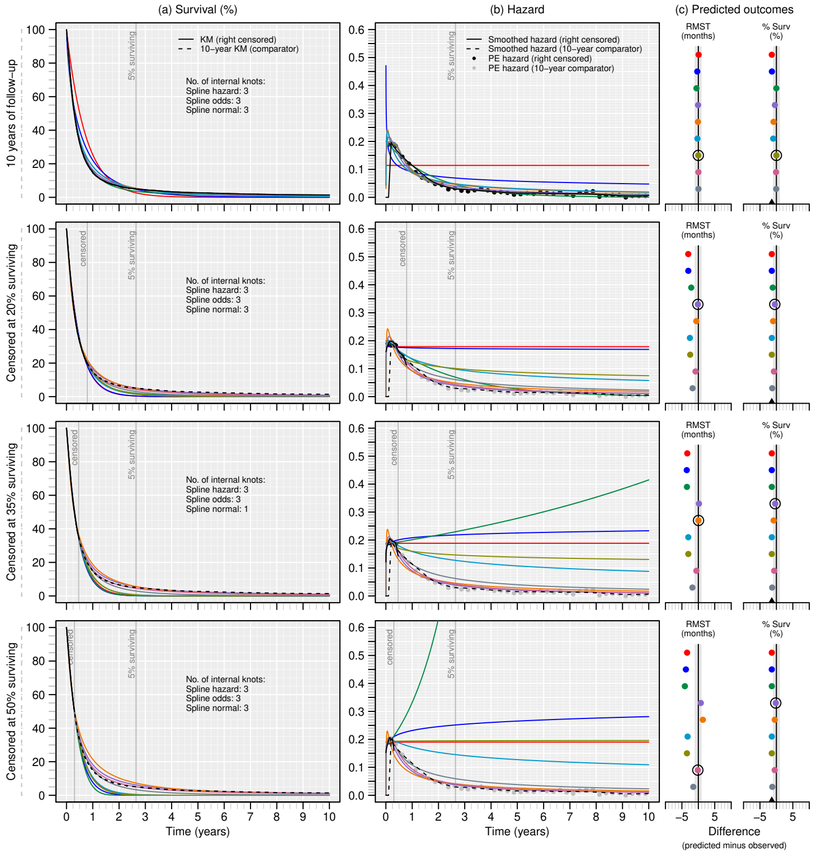
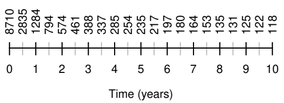


Figure H1. Pancreatic 18–59 years cohort (All models)



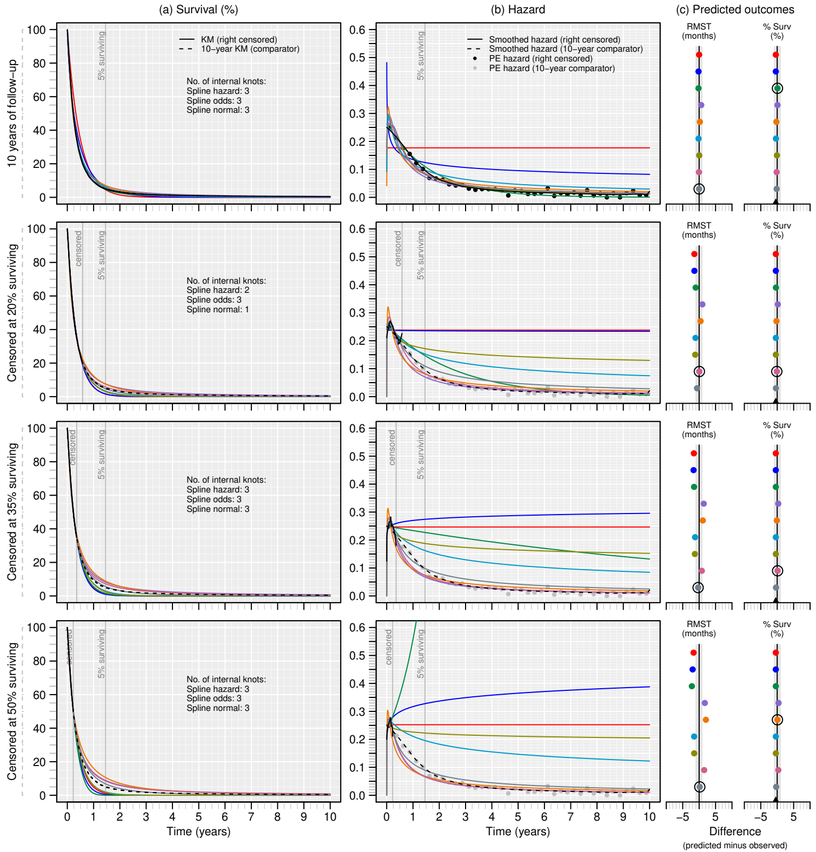
(d) Number at risk



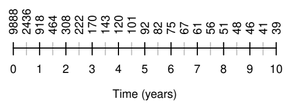
(e) 10–year observed outcomes

RMST = 8.67 months; percentage surviving = 1.40%

Figure H2. Pancreatic 60–69 years cohort (All models)



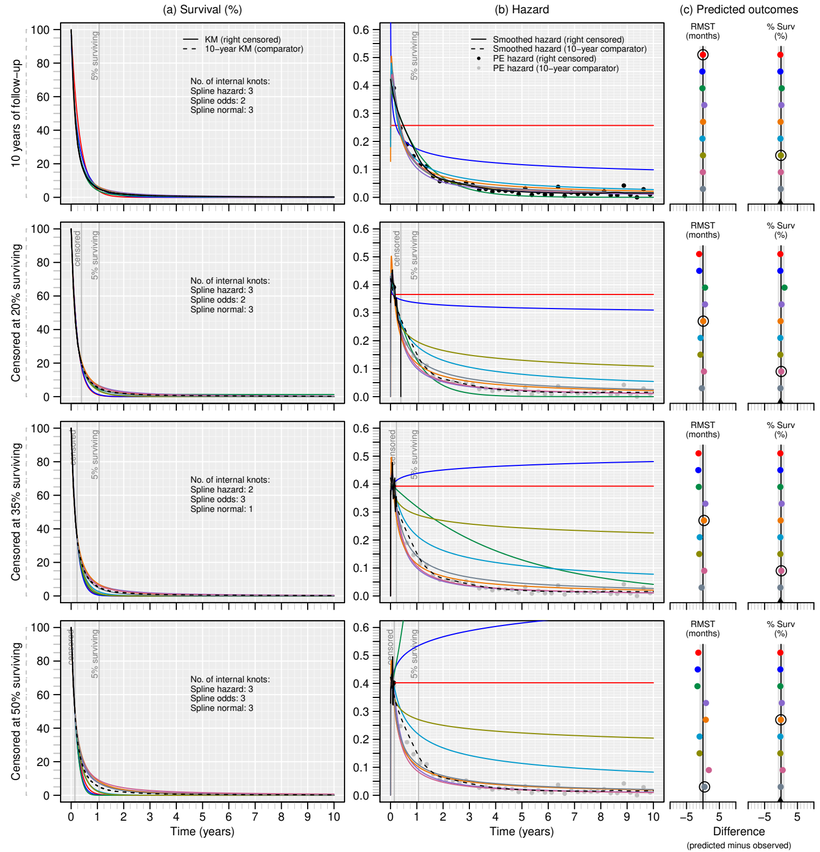
(d) Number at risk



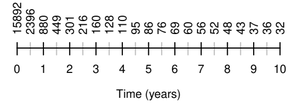
(e) 10–year observed outcomes

RMST = 5.65 months; percentage surviving = 0.43%

Figure H3. Pancreatic 70+ years cohort (All models)



(d) Number at risk



(e) 10–year observed outcomes

RMST = 3.89 months; percentage surviving = 0.21%