WORKPLACE PROJECT PROTFOLIO

Submitted in accordance with the requirements for the Masters of Biostatistics

(Biostatistics Collaboration of Australia)

Project A

A comparison between classic parametric and mixture-cure modelling approaches to estimate observed and predicted survival across various time points

Leonardo Simonella
(SID 199517438)
The University of Sydney
16 December 2016
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model validation</td>
<td>48</td>
</tr>
<tr>
<td>Non-parametric survival outcomes of final XACT trial data-cut</td>
<td>48</td>
</tr>
<tr>
<td>Best-fit parametric survival curves vs observed survival</td>
<td>48</td>
</tr>
<tr>
<td>Extrapolation beyond observed survival</td>
<td>53</td>
</tr>
<tr>
<td>DISCUSSION</td>
<td>56</td>
</tr>
<tr>
<td>APPENDIX</td>
<td>60</td>
</tr>
<tr>
<td>REFERENCES</td>
<td>63</td>
</tr>
</tbody>
</table>
Preface

Project context
In countries such as Australia, Canada and the United Kingdom, government reimbursement of pharmaceutical compounds requires certain HTA guidelines to be met. One of these guidelines includes the criteria of cost-effectiveness of a new drug relative to standard of care. Health economic models used to estimate the incremental cost-effectiveness ratio require lifetime estimates of survival to accurately capture any Quality Adjusted Life Year gain or loss. Clinical trials which are used to inform health economic models have limited observable survival outcomes. This is due to the design of trials needing to satisfy only the scope of regulatory requirements of safety and efficacy. The consequence of this approach leads to clinical trial results to have a number of censored observations. This leads to extrapolated survival estimates beyond the period of observation. Currently, most methods employ classic parametric modelling techniques to estimate the model data generating process, where after a predicted survival curve is used to extrapolate the observed survival trajectory. However for certain contexts, the use of classic parametric survival modelling methods can lead inaccurate estimates of the survival trajectory in patients who are treated with curative intent. To account for deficiencies in current modeling methods, new parametric survival techniques are needed. One such technique is the use of mixture-cure models. In this thesis I explore the use of this technique in the context of a randomized clinical trial in patients diagnosed with stage III colorectal treated with curative intent.

Student’s role: My role in this thesis has been to learn and apply the techniques of parametric survival modelling for both the classic parametric and mixture-cure models. For the classic parametric survival and hazards models, an already built SAS macro was available.
to be used. The macro was developed internally within the MORSE Statistics Unit. I used this macro to run specific parametric survival and hazard functions and curves, but also ran checks (for my own understanding) using the PROC LIFEREG and PROC LIFETEST programs. For the mixture-cure model, a colleague who is part of the MORSE group (HTA Statistician Federrico Felizi) developed the method for Roche HTA submissions. Federrico guided my understanding and assisted by developing a basic model SAS framework based on more complex code developed in R. I used the SAS framework for estimating predicted survival based on a mixture-cure model. Federrico also provided the background survival trajectory for the clinical trial population used in the calculation of the mixture-cure model (This is discussed in detail in the methods section). I lead the statistical analysis, model selection, extrapolation and interpretation of the results.

**Reflections and learnings**

**Communication:** I met regularly with the HTA statistician who developed the mixture-cure model for HTA submissions (Federrico Felizzi). Part of my learning required interpretation and re-communicating the key ideas and techniques to ensure I was able to apply this approach. I also discussed key findings and interpretation of both classic parametric and the mixture-cure models within the context of applying to survival extrapolation with my onsite supervisor (Pierre).

**Work patterns/planning:** Prior to commencing the project I was able to arrange 1 day a week with my workplace supervisor to work on the thesis. He was very supportative, as were my other colleagues, as is the general approach to further education in Roche. Much of the work required my time to be spent on weekends and evenings on developing my knowledge of key concepts and ideas behind classic parametric and mixture-cure survival and hazard functions.
Statistical principles: ‘Prediction’ is a key feature of this thesis. My aim in this project was to derive a valid set of parametric survival models based on an initial set of trial data (data-cut), or training data, then test the best-fit models to determine how well the predicted survival reflects the observed survival from the same trial final data-cut. In contrast to most of the coursework on survival analysis, this project has allowed me to develop an understanding of parametric models and the methods of identifying the data generating process that best fits proposed parametric forms. The limitations of this thesis were the limited quantification of statistical uncertainty and lack of external validity with survival predictions. To date, methods I have used in this project do not fully cover quantification of uncertainty in prediction estimates other than those derived from sampling variability used to construct 95% confidence intervals for non-parametric survival curves. If given more time I would like to have utilized Monte Carlo sampling of estimated variance around the parameter estimates of the parametric functions to explore the range and direction of the predicted survival trajectory. To address the question of external validity, the next step would be to compare the predicted survival estimates with cancer registry data both with and without a similar set of patient characteristics to the clinical trial population.

Statistical methods: To achieve the project objective of identifying the most accurate extrapolated survival trajectory, I applied the methods of both classic parametric and mixture-cure survival modelling. This approach presented a challenge as it was beyond the scope of the general BCA coursework of survival analysis. To undertake the project I needed to extend my knowledge by understanding how parametric and mixture-cure model parameters were derived and apply these methods to derive predicted survival curves. Much of the mixture-cure work was unfamiliar. The general ideas of cure and non-cure in the mixture-cure model
were able to be understood, but the derivation of the parameters required to build a survival model required an understanding of new techniques I was not familiar with. This included the use of SAS code to maximize the log-likelihood to estimate parameters, in particular the step-wise approach to re-estimate parameters based on external data sources (SEER) and applying these to a clinical trial dataset to estimate the predicated and extrapolated survival curves. In addition to the new ideas of mixture-cure models, I needed to develop an understanding of model fitting to observed data as recommended by NICE guidance. In this respect, I learnt about the use and interpretation of the shape of the log-cumulative survival, the empirical estimates of the hazard rates and how they may relate to specific parametric hazard functions, and the use (and limitations) of AIC.

Statistical computing: One of the major challenges in undertaking this project was the use of new computer software and systems. For the BCA coursework, I learnt and applied STATA. However for this project, I needed to use SAS. This was a software I had not used for more than 10 years. In addition, much of the use of SAS involved writing the code in separate software, ‘Nedit’, saving this script, and running data and SAS program from an external server. This was a key challenge that required time to adapt to a new interface and system.

Teamwork

Working with team members: Working with other statisticians was key for this thesis to being completed. Many of the statisticians were able to provide feedback on how certain methods were undertaken, and the reason for their use in the context of cost-effectiveness models. A key area in working with team members was in developing the mixture-cure model. I was able to negotiate specific aspects of the model that I could learn and develop, while my colleague who developed the mixture-cure model for Roche HTA submissions was able to provide data
for aspects related to background mortality (a component of the calculation of predicted survival using the mixture-cure model). This component is part of the calculation to estimate survival for mixture-models. I was also able to obtain feedback from my colleagues regarding notation for the mixture-cure model and interpretation of results.

**Working within timelines:** The main timeline for the project was to complete the project before the due date. Where needed, I was able to devote more time to the project where there was a lull in other aspects of my work. I was able to obtain timely feedback from requests from my supervisor and colleagues. Given the challenging workload for my supervisor and colleagues, I had to adapt feedback for specific analyses at selected timepoints that suited their needs. Hence flexibility was a skill I needed to develop for the project to be complete.

**Ethical considerations**

The clinical trial data used in this thesis was approved by FDA and EMA at the time of regulatory approval. The key ethical consideration for this project was to follow scientific principle of methodological validity. My approach was to develop a best fit parametric model based on the trial’s initial datacut. To ensure my method was valid, I did not re-estimate the prediction based on further assessment of the final data-cut. This would be unethical and undermine the validity of the approach. In writing up the results, I undertook the approach that, whether a good fit or not, the pros and cons of the best selected model should be discussed relative to other parametric survival model options.
Declaration

Declaration by Student

I declare that this project is evidence of my own work, with direction and assistance provided by my project supervisor, Pierre Ducournau, and workplace colleague Federrico Felizzi. This work has not been previously submitted for academic credit.

……………………………………………
Leonardo Simonella

……………………………………………
Date

Comments and Declaration by Project Supervisor

The declared contributions by this student are true and correct. To my knowledge, the involvement and effort of this student for this project is satisfactory for the requirements of a BCA Workplace Project.

……………………………………………
Pierre Ducournau

……………………………………………
Date
**Introduction**

One of the key features of the Health Technology Assessment (HTA) framework is the need to include a cost-effectiveness analysis (CEA) of a new intervention [1]. This type of evaluation compares the added benefit and associated costs of a new intervention relative to the current standard of care. To fulfill a CEA one needs estimates around the costs and benefits of the interventions being compared. More often than not, the costs and benefits are assessed across a patient’s lifetime [2]. However, in the context of clinical trials, the resource requirements for long term evaluation of costs and benefits are prohibitive and not required by regulatory authorities. Consequently, to satisfy the needs of HTA bodies, survival outcomes modelled in CEA need to be extrapolated beyond observed survival (observed in clinical trials).

In general, the extrapolation of survival outcomes is based on fitting an appropriate parametric survival model to the observed clinical trial data [2]. The choice of parametric models used to fit the distribution around a particular set of survival data can include an exponential, Weibull, log-logistic model, Gompertz, or a log-normal model [3]. However, when imposing one of these model structures onto a set of survival data, certain criteria need to be used to evaluate the fit of the model to the observed data. These assessments include statistical testing (AIC, BIC, sum of squared deviations, Martigale residuals), visual inspection, clinical validity, and the potential to use external data (if available) [4]. However, even with these various assessments, there is the possibility that the ‘best’ fit model does not plausibly follow the long term observed outcomes [5].
Mixture-cure models provide an alternative approach to the classic set of parametric models [6]. In certain treatment contexts classic parametric models may not accurately reflect survival outcomes over longer observation periods. Such treatment contexts include adjuvant treatment where there is curative intent. For patients diagnosed with non-metastatic staged disease, often the choice of treatment involves excision or organ removal, followed by adjuvant chemotherapy and or radiotherapy. For some patients, the residual cancers may not be eliminated, and hence will likely follow an accelerated failure time, which may be captured accurately using standard parametric models. However, for other patients where the disease is cured, standard parametric models may overestimate the rate of failure over longer periods of observation. Consequently, patients who fall into the cured category will follow a survival trajectory that follows the population background mortality. Mixture-cure survival models provide a platform that takes into account the survival trajectory of both cured and non-cured population.

One context in which mixture-cure survival models can be compared with standard parametric survival models is in the randomized clinical trial setting. Large scale randomized clinical trials which evaluate the safety and efficacy of pharmacological compounds have several points of ‘inspection’ of clinical trial data (datacuts) by a data monitoring committee. The initial datacut provides an opportunity to assess which parametric model fits best to the data generating process. At this initial stage there are a limited number of outcomes due to the shorter period of observation. At the final datacut there are likely to be more observed outcomes. It is the latter data set that provides an opportunity to assess the predictive fit of the initial standard parametric or mixture-cure model.
The XACT study is a randomized clinical trial designed to compare the safety and efficacy of adjuvant oral fluoropyrimidine capecitabine to that of bolus 5-fluorouracil and leucovorin (5-FU/LV) in patients diagnosed with stage III colon cancer [7]. 1987 patients were enrolled in 164 centres around the world between 1999-2001. The primary and secondary endpoints of the study were disease free (DFS) and overall survival. The study was designed to assess the equivalence of DFS of capecitabine to that of bolus 5-FU/LV. In the cost-effectiveness analysis used for the HTA submission for drug reimbursement, survival outcomes associated with capecitabine were based on an initial data-cut of the trial extrapolated using certain parametric survival assumptions. However, to date it remains uncertain as to how well the parametric survival curves fitted to observe survival outcomes over the longer term.

My aims in this thesis were therefore to:

1. Fit a set of parametric and mixture cure models to overall survival in patients diagnosed with stage III colorectal cancer treated with curative intent, based on an initial data-cut of the XACT clinical trial.

2. Assess which set of parametric models (standard and mixture-cure) best fits the overall survival, based on an initial data-cut of the XACT clinical trial.

3. Evaluate how well the selected best fitting parametric models predicts observed overall survival based on the final data-cut of the XACT clinical trial.

4. Extrapolate the best fit parametric and mixture cure survival model beyond the observation period, over the long-term.
Materials and Methods

Data source
The XACT trial is a randomized clinical trial with several datacuts. All data had been cleaned, so no additional data cleaning was necessary. The first datacut on outcomes (file: i66001d) was made available based on a median follow-up time of 2.98 years (min: 0.003 yrs – max: 5.41 years), while the final datacut on outcomes (file: i66001h) was made available after a median follow-up time of 5.20 years (min: 0.003 yrs – max: 8.63 years). For each datacut, patient outcomes were recorded as having one of the following events: progressive disease (relapse of recurrence), death, or no event. A ‘progression’ or ‘death’ event were recorded as a ‘1’, otherwise patient events were censored ‘0’ (no event). For a particular datacut, a specific point in time was used define the time-to-event or censoring (no event during time follow-up interval). Therefore a failure indicator for person ‘i’ was defined as $\delta_i = 1$ if $T_i \leq U_i$ or $\delta_i = 0$ if $T_i > U_i$, where T and U are time-to-event and observation time, respectively. Observation time referring to date of initiation of treatment until last date of observation. Similarly, a censoring indicator for person ‘i’ was defined as $c_i = 0$ if $T_i \leq U_i$ or $c_i = 1$ if $T_i > U_i$. All statistical analysis was conducted using SAS 9.4.

Parametric survival models
Parametric survival models can be defined as mathematical expressions of time-to-event phenomena that follow a particular probability distribution. These include exponential, log-normal, log-logistic, gamma, to name a few [3, 8]. The mathematical expression for these distributions need to be compared with observed time-to-event outcomes to determine which distribution is likely to be reflected in the observed data. The expressions for distributions used for comparing observed time-to-event outcomes include the survival function $S(t)$, the
hazard function $h(t)$, and the cumulative hazard function $H(t)$. The survival function is expressed as follows:

$$S(t) = \Pr(T > t)$$

$$= \int_t^\infty f(\mu) d\mu$$

$$= 1 - F(t)$$

The definition of the survival function is the probability the random variable $T$ (a random variable that can take on any number greater than or equal to 0) is greater than some the observed value $t$. The function $f(\mu)$ represents the probability density function of the random variable, such that the probability of a specific value of the random variable $T$ is $f(t)$. The function $F(t)$ is the cumulative probability of failure, such that $F(t) = \Pr(T \leq t)$.

The hazard function, defined as the instantaneous rate of failure, is also called a conditional failure rate. This is because the failure rate is based on the probability of failure in a particular time interval given one has not had a failure before the beginning of the time interval. This can be expressed as:

$$h(t) = \lim_{\Delta t \to 0} \frac{\Pr([t + \Delta t] > T > [t|T > t])}{\Delta t}$$

$$= \frac{f(t)}{S(t)}$$

Finally, the cumulative hazard function can be defined as the number of failure times on average one would observe during a particular interval. This is expressed as:
\[ H(t) = -\ln[S(t)] \]

**Modelling parametric survival models to XACT clinical trial data**

In this analysis I utilized a pre-programmed SAS macro to identify the best fitting parametric survival curve for each datacut of the XACT clinical trial. The macro presented in the appendix was developed by the MORSE team at Hoffman-La Roche’s. For the purposes of the macro the clinical trial data is divided into two separate tables, one reflects participants’ demographic details (fixed) and the other reflects efficacy, or outcomes, of participants (datacuts). It is the efficacy dataset that is updated. The first datacut (file: i66001d) is used to estimate the best fitting parametric curve for the data generating process. The final datacut (file: i66001h) is used as the ‘gold standard’ to compare how well the best fitting parametric survival predicts observed outcomes (death).

Built into the macro are a series of parametric and non-parametric functions. For all analyses I define the covariate \( x \) as a dichotomous variable for whether a participant received the control drug (5U+Leucovorin) \( x = 0 \), or the intervention drug (Capecitabine) \( x = 1 \). The first parametric function that is modelled on the observed data is based on the exponential distribution. For each treatment arm the hazard and survival functions are calculated as:

\[
\begin{align*}
    h(t_i|x = 1) &= \exp(-[\beta_0 + \beta_1 x]) \\
    S(t_i|x = 1) &= \exp[-\exp(-[\beta_0 + \beta_1 x]) * t_i] \\
    h(t_i|x = 0) &= \exp(-\beta_0) \\
    S(t_i|x = 0) &= \exp[-\exp(-\beta_0) * t_i]
\end{align*}
\]
Where $\beta_0$ is the intercept (the coefficient value of the control arm, 5U+Leucovorin) while the $\beta_1$ is the coefficient value for the effect of the intervention arm (Capecitabime) on death. For data generating process that follow a **Weibull distribution**, the hazards and survival functions are expressed as:

\[
h(t_i | x = 1) = \left( \frac{1}{p} \right) \exp \left( - \frac{(\beta_0 + \beta_1 x)}{p} \right) \ast t_i^{(\frac{1}{p} - 1)}
\]

\[
S(t_i | x = 1) = \exp[ - \exp \left( - \frac{(\beta_0 + \beta_1 x)}{p} \right) \ast t_i^{\frac{1}{p}} ]
\]

\[
h(t_i | x = 0) = \left( \frac{1}{p} \right) \exp \left( - \frac{\beta_0}{p} \right) \ast t_i^{(\frac{1}{p} - 1)}
\]

\[
S(t_i | x = 0) = \exp[ - \exp \left( - \frac{\beta_0}{p} \right) \ast t_i^{\frac{1}{p}} ]
\]

The additional parameter $p$ is the scale parameter, such that values of the scale parameter determine the hazard function shape. If the scale is $p = 1$, then the above expression reduces to the exponential distribution described above. A data generating process for the **log-normal distribution** has the following hazard and survival function characteristics for each treatment arm:

\[
h(t_i | x = 1) = \frac{\varphi \ast \left( \frac{\log(t_i) - (\beta_0 + \beta_1 x)}{p} \right)}{t_j - (t_j \ast \varphi) \ast \left( \frac{\log(t_i) - (\beta_0 + \beta_1 x)}{p} \right)}
\]
\[ S(t_i | x = 1) = 1 - \varphi \left( \frac{[\log(t_i) - (\beta_0 + \beta_1 x)]}{p} \right) \]

\[ h(t_i | x = 0) = \frac{\varphi \left( \frac{[\log(t_i) - \beta_0]}{p} \right)}{t_j - (t_j \times \varphi) \times \left( \frac{[\log(t_i) - \beta_0]}{p} \right)} \]

\[ S(t_i | x = 0) = 1 - \varphi \left( \frac{[\log(t_i) - \beta_0]}{p} \right) \]

where the \( \varphi \) is the cumulative distribution function of the standard normal distribution. The data generating process for the log-logistic distribution has the following hazard and survival function:

\[ h(t_i | x = 1) = \left( \frac{1}{p} \exp \left( -\frac{[\beta_0 + \beta_1 x]}{p} \right) \times t_i^{\left( \frac{1}{p} - 1 \right)} \right) \left( 1 + \exp \left( -\frac{[\beta_0 + \beta_1 x]}{p} \right) \times t_i^{\frac{1}{p}} \right) \]

\[ S(t_i | x = 1) = \frac{1}{\left( 1 + \exp \left( -\frac{[\beta_0 + \beta_1 x]}{p} \right) \times t_i^{\frac{1}{p}} \right)} \]

\[ h(t_i | x = 0) = \left( \frac{1}{p} \exp \left( -\frac{-\beta_0}{p} \right) \times t_i^{\left( \frac{1}{p} - 1 \right)} \right) \left( 1 + \exp \left( -\frac{-\beta_0}{p} \right) \times t_i^{\frac{1}{p}} \right) \]
\[ S(t_i|x = 0) = \frac{1}{1 + \exp \left( \frac{-\beta_0}{p} \right) \cdot t_i^\frac{1}{p}} \]

These specified functions have been adapted from an internal Roche document title ‘Stem Macro Derivations’. The macro is based on pre-specified SAS procedural codes including PROC LIFREG, and PROC LIFETEST.

**General criteria to assess parametric survival curve fit**

To date no fixed standard exists to determine the best fitting parametric curve. The National Institute for Health and Clinical Excellence (NICE) has provided a technical report to guide economic evaluations submissions of health technology interventions that require extrapolated survival curves for clinical trial patient level data [2].

The key approaches outlined by NICE to guide assessment of suitability of survival models include visual inspection, log-cumulative hazard plots, and AIC/BIC tests. The visual inspection approach entails exploring how well a parametric curve follows a non-parametric (Kaplan-Meier) survival curve of the time-to-event data. However, caution needs to be taken with this approach where there are points of heavy censoring along certain parts of the curve, in particular the tail ends. The parametric curve may be appropriate for certain segments, but not for others.

Log-cumulative hazard plots provide insight into how the hazard occurs over time. Different parametric functions described above present hazards that can be either constant (exponential), monotonic (Weibull, Gompertz) or non-monotonic (log-logistic, log-normal, gamma). The use of a log-cumulative hazard, which is based on a log of the –log of the
survival function over the log of time can assist in describing the nature of the hazards. The changing nature of the gradient of the log-cumulative hazard curve can assist in deciphering the nature of the hazard.

An alternative approach to estimate the hazard shape is to plot an estimate of the hazard using a kernel smoothing estimator. The approach applies a kernel weight to the differences in the cumulative hazard estimate between event times. An approximation for the cumulative estimate is the Nelson-Aalen estimator:

\[ \hat{H}(t) = \sum_{j \mid t_j < t} \frac{d_j}{n_j} \]

Where \( d_j \) is the number of failures at time \( j \), and \( n_j \) is the number at risk at \( j \), where the sum of each event at \( t_j \) is up to \( t \). The estimate of the hazard function at each event time is:

\[ \hat{h}(t) = b^{-1} \sum_{j=1}^{D} K_t \left( \frac{t - t_j}{b} \right) \Delta \hat{H}(t_j) \]

Where \( b \) is the bandwidth, \( D \) is the total number of event times \( j = 1 \ldots D \), \( K_t \) is some kernel function and \( \Delta \hat{H} = \Delta \hat{H}(t_j) - \Delta \hat{H}(t_{j-1}) \).

However, caution needs to be taken with this approach with different kernel smoothing estimators and their associated bandwidths. Broadly speaking, too broad a bandwidth used in a kernel estimator can lead to bias in the interpretation of the curve (underfitting), whereas too narrow bandwidth can lead to extreme variance (overfitting) displaying a more jagged curve. In the Roche macro, the choice of macros are Uniform, Epanechnikov, and biweight.
kernel smooth estimators. The Uniform kernel is defined as $K(u) = 1/2$, the Epanechnikov is defined as $K(u) = \left(\frac{3}{4}\right) \ast (1 - u^2)$, while the Biweight is defined as $K(u) = \frac{15}{16} \ast (1 - x^2)^2$. Where $u = \left(\frac{t - t_i}{b}\right)$ in this context. The choice of bandwidth is based on a SAS preprogrammed golden section search algorithm. The algorithm identifies a bandwidth that minimizes the mean integrated squared error for the event times. This is defined as:

$$MISE(b) = \mathbb{E}\left[ \int_{\tau_L}^{\tau_U} \left( \hat{h}(t, b) - \tilde{h}(t) \right)^2 \right]$$

Where the expectation is over various values of $h(t)$ integrated at selected intervals of times of events ($\tau_L$ to $\tau_U$).

The final key approach that is presented for assessing the fit of a parametric survival model is the use of Akaike’s Information Criteria (AIC) and the Bayesian Information Criteria (BIC). The calculation of these estimates are:

$$AIC = 2K - 2\ln(L)$$

$$BIC = -2\ln(L) + K \ast \ln(n)$$

Where $K$ is the number of parameters used in the parametric equation, $L$ is the maximized value of the likelihood, and $n$ is the sample size. The key idea behind these approaches is that the more parameters that are used to fit a parametric curve to the observed data, there is a risk that there may be overfitting. The resulting overfit prevents an understanding of the true data
pattern that is being generated. Thus the AIC and BIC provide scores that penalize parametric models that have more parameters. In the context of parametric survival models, Weibull and Log-logistic models will have a higher penalty than the exponential model. The interpretation of an AIC or BIC score is that, when comparing parametric models, the lower the AIC or BIC value the better the model fit. Therefore, for parametric models with more parameters to have a better fit, the sums of square error need to be far smaller when extra parameters are used.

Cure survival model

A cure survival model assumes that certain individuals do not succumb to an event of interest \cite{6}. In the context of a cancer diagnosis, these individuals do not succumb to death due to the diagnosed cancer, but rather to other factors. These persons are assumed to be ‘statistically cured’, and thus are assumed to follow a mortality risk trajectory which follows the background population mortality risk. Several types of cure survival models are discussed in Lambert et al. However, for the purpose of this thesis, I will focus on the mixture-cure model that has been amended and developed by the Roche MORSE Statistics group in the Global Pricing and Market Access division.

The general mixture-cure survival and hazard functions are specified as:

\[
S(t) = S^\pi(t) \ast [\pi + (1 - \pi) \ast S_u(t)]
\]

\[
h(t) = h^\pi(t) + \left[\frac{(1 - \pi) \ast f_u(t)}{\pi + (1 - \pi) \ast S_u(t)}\right]
\]
The survival function $S^#(t)$ can be broadly defined as an averaged background population-based survival outcome of the study sample conditioned on the study sample participant’s age, country, or other characteristic that differentiate survival outcomes. The hazard function $h^#(t)$ is the instantaneous mortality risk at a particular age, given age, country or other characteristics. Both $S^#(t)$ and $h^#(t)$ are derived from country specific life-tables. The parameter $\pi$ is a cure fraction, which is defined as the estimated proportion of the sample assumed to not succumb to the event of interest. The final two parameters are the probability density function $f_\alpha(t)$ and survival function $S_\alpha(t)$ for the uncured population of the sample. These are functions that are fitted according to best fitting parametric function criteria for the time-to-event data generating process. The process for determining the best fit has been described previously.

**Mixture-cure model for stage III colorectal cancer**

The steps I used to calculate the overall cure survival function for the XACT clinical trial participants were as follows.

**Step 1: Data sources to estimate the cure fraction**: The cure fraction is estimated using a combination of background population-based mortality risk and a disease specific registry. In this analysis I used the US specific lifetables to estimate background mortality and the SEER cancer registry data to represent the disease specific registry. The SEER cancer registry is population-based and covers 28% of the population through 18 cancer registries throughout the United States [9]. The choice of US data to estimate the cure fraction was based on the availability of the SEER cancer registry data. Other population-based cancer registry sources could also be used.
From the SEER registry I extracted all patients diagnosed with colorectal cancer between Jan-1998 to Dec-2001 with stage III disease. This resulted in n=4302 cases. After excluding those with a recorded 0 survival time, 4172 cases remained in the dataset. Variables extracted for each case included age and year at diagnosis, survival time, censoring status and stage of disease. Background population-based mortality risk were used in the form of US specific lifetables that were age and gender specific for the years 1933 to 2014 [10].

To undertake the estimation process for the cure fraction, a merged dataset was needed that included the probability of survival for the disease of interest, and the background mortality risk. In this case, each unit record of stage III colorectal cancer was merged with mortality risk based on US specific lifetables. US specific lifetables were merged to each individual in SEER based on sex, age at death and year at death. In the US lifetables, $nQ_x$ represents the probability of death for a particular age in a one year interval. This is a proxy for instantaneous background mortality risk. For the merge, the US lifetable variables ‘age’ and ‘year’ were re-coded to ‘age at death’ and ‘year at death’. In the extracted SEER dataset ‘age at death’ was defined as ‘age at diagnosis’ plus ‘survival time’, while ‘year of death’ was defined as ‘year of diagnosis’ plus ‘survival time’. The resulting merged table gives each participant with stage III colorectal cancer diagnosed in 1998-2001 with an assumed population-based background instantaneous risk of death, had they not had the disease of interest (in this case colorectal cancer). This estimate is used in the subsequent step to estimate the cure fraction for stage III colorectal cancer for patients treated in this period.

**Step 2: Estimation of cure-fraction:** The merged dataset between the SEER colorectal stage III patients and US lifetables was used to estimate the cure fraction based on the method of maximum likelihood. The log-likelihood function evaluated was:
\[
\log L_{\text{seer}}(\theta, \pi|t, \delta) = \sum_{i=1}^{N} \delta_i \log[h_{\text{mixed \ cure}}(t_i|\theta_{1}^{\#}, \pi^{\#})] + \sum_{i=1}^{N} \log[\pi + (1 - \pi) \ast S_{\text{uncured}}(t_i|\theta_{1}^{\#}, \pi^{\#})]
\]

Where \( N \) is the total number of SEER stage III colorectal cancer patients \( (i = 1, 2, \ldots, 4172) \), and \( \delta_i \) is the censoring variable, where a 1= event (death) and 0=censored. In the function, \( \log h_{\text{mixed \ cure}} \), \( t_i \) is time-to-event in years for person \( i \), \( \theta_{1}^{\#}, \pi^{\#} \) are initial parameter estimates that act as seeding values to be re-estimated as \( \theta \) and \( \pi \) using the Newton-Raphson root variable approximation. The method of approximating the maximum likelihood for each parameter was based on the PROC NLMIXED function (discussed further below).

The hazard function in the log-likelihood equation is calculated as:

\[
h_{\text{mixed \ cure}}(t_i|\theta_{1}^{\#}, \pi^{\#}) = nQx_j + \frac{(1 - \pi) \ast f_{\text{uncured}}(t_i|\theta_{1}^{\#})}{(\pi + 1 - \pi) \ast S_{\text{uncured}}(t_i|\theta_{1}^{\#})}
\]

The term \( nQx_j \) is a fixed estimate of the \( j \)th interval instantaneous background mortality for person \( i \) in the SEER database had they not been diagnosed with stage III colorectal cancer. The \( j \)th interval for the instantaneous mortality risk is allocated to each person based on the addition of their age and year at diagnosis plus the time to their event (i.e. death). For example, the background population instantaneous mortality risk in the US in the year 2004 for a female at age 68 is 0.01468. This risk would be allocated to any female in the SEER database diagnosed with colorectal cancer female who had an event at aged 68 yrs and their death was in the year 2004. The diagnosis would have to been in the year 2002 at age 66 years.
The other components of the hazard function are the probability density function 
$f_{uncured}(t_i | \theta^{#1})$ and survival function $S_{uncured}(t_i | \theta^{#1})$ for the uncured population. These functions are estimated for person $i$ based on the vector of parameters $\theta^{#1}$. The parameters are based on the best fitting parametric function for the time-to-event data for SEER patients diagnosed with stage III colorectal cancer. The basis for determining the best fitting parametric curve has been described previously.

To approximate the maximum likelihood I used the PROC NLMIXED procedure. The procedure code to evaluate the log-likelihood parameters of interest is presented below:

```plaintext
Line 1 - proc nlmixed data=colrect_only_d;
Line 2 - parms / data = parm_surv_estimates;
Line 3 - f_t = pdf('parametric_dist.', ttdied_yrs, parameter_1, 
parameter_2, parameter_x, ...);
Line 4 - S_t = 1 - cdf('parametric_dist.', ttdied_yrs, parameter_1, 
parameter_2, parameter_x, ...);
Line 5 - h_ = qx + (1-pi_)*f_t/(pi_ + (1-pi_)*S_t);
Line 6 - s_ = pi_ + (1 - pi_)*S_t;
Line 7 - ret_val = cens*log(h_) + log(s_);
Line 8 - model ttdied_yrs ~ general(ret_val);
Line 9 - ods output parameterestimates = seer_pop_cure_fraction;
run;
```

The first line of the PROC NLMIXED code reflects the dataset used to evaluate the log-likelihood, in this case I used the merged SEER and US lifetable dataset. The second line is the dataset containing the vector of parameters for the best fitting parametric survival distribution to the SEER time-to-event dataset. For example, if the log-logistic function was the best fitting parametric model, then the intercept, scale parameter are listed. Lines three and four are the probability density functions $[f_{uncured}(t_i | \theta^{#1})]$ and the survival functions $[S_{uncured}(t_i | \theta^{#1})]$ for person $i$, where the calculation for each function is determined by specifying the selected parametric survival function equation, the outcome variable (i.e. survival time), and the parameters used to calculate the parametric survival function (parameter 1, parameter 2, parameter x, ...). Lines five and six calculate the mixed cure
hazard function $h_{\text{mixed\_cure}}(t_i|\theta^1, \pi^#)$, and the second component of the log-likelihood equation: $\sum_{i=1}^{N} \log[\pi + (1 - \pi) * S_{\text{uncured}}(t_i|\theta^1)]$. In line five $q_X$ refers to $nQx_j$. Line seven is the calculation of the complete log-likelihood for each person $i$ in the merged SEER dataset. Line eight runs the Newton-Raphson optimization algorithm to estimate the maximum likelihood for the parameters in the log-likelihood function. Finally, line nine outputs the unknown parameters estimated through the maximum likelihood. These include the cure fraction, and re-estimated initial parameters indicated in line 2.

Step 3: Estimation of parametric survival curve parameters in the context of a mixture-cure model:

Following an estimate of the cure fraction, I next estimate the mixture-cure model parameters for each treatment arm in the Xeloda clinical trial. Similar to the merged SEER and US dataset, I merged each observation in the Xeloda clinical trial dataset with country specific lifetables according to sex, country, age at death and year of death.

In the first instance, initial parametric survival distribution parameters are estimated for each treatment arm based on the assumed context that there is no cure. These initial starting parameters are substituted in the log-likelihood equation of each treatment arm, along with the derived fixed estimate of the cure fraction:

$$\log L_{\text{SFU+Lex}}(\theta^*|t, \delta, \pi) = \sum_{i=1}^{N} \delta_i \log[h_{\text{mixed\_cure}}(t_i|\theta^{#2}, \pi)] + \sum_{i=1}^{N} \log[\pi + (1 - \pi) * S_{\text{uncured}}(t_i|\theta^{#2})]$$

$$\log L_{\text{Cap}}(\theta^*|t, \delta, \pi) = \sum_{i=1}^{N} \delta_i \log[h_{\text{mixed\_cure}}(t_i|\theta^{#2}, \pi)] + \sum_{i=1}^{N} \log[\pi + (1 - \pi) * S_{\text{uncured}}(t_i|\theta^{#2})]$$
Where $\theta^*$ are the re-estimated vector of parameters for the parametric survival distribution that has an assumed cure fraction. The vector of parameters $\theta^2$ are initial seeding values based on the best fitting parametric survival distribution in the assumed context that there is no cure. Similar to the The log-likelihood function for the SEER dataset, the log-likelihood for each treatment arm is evaluated using method of maximum likelihood using the PROC NLMIXED procedure.

**Step 4: Calculation of mixture-cure survival curves:** Once parameters for the mixture survival model are calculated, the next step is substitution of estimated parameters into the survival functions:

$$S_{SFU+Lev}(t_i) = S^#(t)_j \cdot [+(1-\pi) \cdot S_u(t_i|\theta^*)]$$

$$S_{Cap}(t_i) = S^#(t)_j \cdot [\pi + (1-\pi) \cdot S_u(t_i|\theta^*)]$$

Where $S^#(t)_j$ is an ‘averaged’ background population survival probability for a particular follow-up year (jth group, where $j = year\ 1, year\ 2 \ldots year\ 30$). $S^#(t)_j = \frac{1}{N} \sum_{i=1}^{N} \frac{S_B(a_i+t)}{S_B(a_i)}$.

Where $\frac{S_B(a_i+t)}{S_B(a_i)}$ is a conditional probability defined as a clinical trial participant’s ‘theoretical’ non-cancer survival probability for their gender country of treatment ($S_B$) at age ($a_i$) plus ($t$), where $a_i$ is the age at the point of randomization and $t$ is the time to the event (death), divided by trial participant’s ‘theoretical’ non-cancer survival probability ($S_B$) at age ($a_i$). The summation is over all the individuals in the clinical trial. Figure 1 displays the background survival probability of individual trial participants had they not had the disease
(black survival probability curves lines) and the average for the sample (red survival probability)

The remaining parameters for the cure model’s overall survival function are the cure fraction \((\pi)\) and the re-estimated vector of parameters for the best fitting parametric survival population for the non-cure group \(S_u(t_i|\theta^*)\). In this case \(\theta^*\) is the re-estimated vector of parameters based on the notion that a cure proportion has been accounted for in the maximization of the log-likelihood.

**Figure 1**: XACT trial participant’s background survival probability of stage III colorectal cancer patients (black) and their average background survival probability for all trial participants (red)

Step 5: Fitting mixture-cure survival curves: - The final step in estimating survival with the use of the mixture-cure model was re-calibrating the parameters to fit the observed data. As outlined previously, the derivation of the cure fraction for the clinical trial population with
stage III colorectal was derived from an external source population. In this case, the cure fraction was derived from the US SEER population of stage III colorectal cancer who were diagnosed in the same time period as the participants in the XACT clinical trial. One of the concerns with this approach is that the external population might not reflect the broad characteristics of the trial population. For example, the population used to derive the cure fraction may be older, have more advanced disease or some other characteristic of disease that leads to a greater accelerated failure time. This can be observed by comparing the non-parametric curves for the clinical trial population (based on the first data cut) and the SEER population (Figures 1 and 16). The survival curves show that the clinical trial population has a higher survival probability at the 4-5 year follow-up period (60-80%) compared with the SEER stage III population (50-60%). Therefore, to adjust for the lower survival probability from which the cure fraction was derived, the cure fraction may need to be re-calibrated. In this case the cure fraction needs to be re-calibrated upwards to a point where the mixture-cure survival estimate follows the general trend of the clinical trial population. The approach used to fit the re-calibrated curves was based a visual comparison of multiple cure fraction estimates to the observed non-parametric curve for each treatment arm.

**Extrapolation beyond observed survival:**

Each treatment arm had survival extrapolated beyond observed survival time. This was based on using the selected best fit parametric survival curve and the re-calibrated mixture-cure survival model (as defined above), and applying the model’s parameters to time (years) beyond what was observed in the clinical trial. For the 5FU+Leucovorin arm extrapolation was from the last participant observation time of 8.63 years to 50 years, while for the Capecitabine it was from 8.54 years up to 50 years as well.
Results

\textit{XACT trial}

In the XACT trial there were 1987 participants across 25 countries diagnosed with stage III colorectal cancer. 983 were randomized to 5FU+Leucovorin, while 1004 were randomized to Capecitabine. The median ages at diagnosis of participants in each arm were 63 and 62 years, respectively. Both treatment arms have left skewed age distributions, with most patients (85%) in the 55 to 85 year age group (Figure 2). The remaining 15% of participants were aged between 20 to 55 years. Most participants were diagnosed with stage IIIB disease (62% and 60%), while in both treatment arms there were more males (54%) than females (46%).

In the first datacut (i66001d) 23% of participants in the 5FU+Leucovorin arm had an event (died), while in the Capeciabine arm it was 20%. In the final datacut (i66001h) the proportion of patients who had an event increased to 36% and 32%. The median follow-up time for the first datacut was approximately 3 years (maximum follow-up 5 years), while in the final datacut it was around 5 years (maximum follow-up 8 years).

Non-parametric survival outcomes: In the first datacut median survival was not reached in either treatment arm (Figure 3). Among those randomized to 5FU+Leucovorin 25% of participants had an event at 3.45 years (95% CI 3.06 – 4.11 years). For those randomized to Capecitabine arm it was 3.98 yrs (95% CI 3.68 - .). However this difference was not statistically significant (Log-rank 3.23, p-value 0.07; Wilcoxon test value 3.08, p-value 0.079).
Table 1: Baseline characteristics of clinical trial population by treatment arm

<table>
<thead>
<tr>
<th></th>
<th>5FU + Leucovorin n (%)</th>
<th>Capecitabine n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Number at randomization (ITI)</strong></td>
<td>983</td>
<td>1004</td>
</tr>
<tr>
<td><strong>Number of events</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>First data cut (i6600d)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Event (death)</em></td>
<td>227 (23)</td>
<td>200 (20)</td>
</tr>
<tr>
<td><em>Censored</em></td>
<td>756 (77)</td>
<td>804 (80)</td>
</tr>
<tr>
<td>Final data cut (i6600h)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Event (death)</em></td>
<td>351 (36)</td>
<td>319 (32)</td>
</tr>
<tr>
<td><em>Censored</em></td>
<td>632 (64)</td>
<td>685 (68)</td>
</tr>
<tr>
<td><strong>Median follow-up</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>First data cut (i6600d)</td>
<td>2.95 yrs (min: 0.03 – max: 5.42)</td>
<td>3.01 yrs (min: 0.03 – max: 5.23)</td>
</tr>
<tr>
<td>Final data cut (i6600h)</td>
<td>5.12 yrs (min: 0.03 – max: 8.63)</td>
<td>5.27 yrs (min: 0.03 – max: 8.54)</td>
</tr>
<tr>
<td><strong>Sex</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>532 (54)</td>
<td>542 (54)</td>
</tr>
<tr>
<td>Female</td>
<td>451 (46)</td>
<td>461 (46)</td>
</tr>
<tr>
<td><strong>Median age (range)</strong></td>
<td>63 yrs (Q1: 55, Q3:69)</td>
<td>62 yrs (Q1: 54, Q3:68)</td>
</tr>
<tr>
<td><strong>Tumor stage</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stage IIIA</td>
<td>82 (9)</td>
<td>92 (9)</td>
</tr>
<tr>
<td>Stage IIIB</td>
<td>612 (62)</td>
<td>603 (60)</td>
</tr>
<tr>
<td>Stage IIIC</td>
<td>288 (29)</td>
<td>305 (31)</td>
</tr>
</tbody>
</table>

Figure 4 presents the estimated observed hazard rate for the first data cut. The rate at which events occur in each treatment arm appears to increase until approximately 2-2.5 years. After this time point, the rate of events appears to remain constant in the Capecitabine arm, whereas the rate of events in the 5FU+Leucovorin arm continues to increase at a decreasing rate. Figures 5 and 6 provide additional insight into the rate of events. Figure 5 suggests a steady similar rate from beginning in each treatment arm until approximately 2 years. After this time point, participants in the 5FU+Leucovorin arm display an increasing rate of events relative to the Capecitabine arm. Figure 6 shows a similar pattern with a slower rate of events initially, then an increase in the rate with a slight divergence at 2.5 yrs (ln(0.92) on the x-axis of figure 6).
Figure 2: Histogram of age at randomization overlay by treatment group

The figure shows a histogram of age at randomization overlay by treatment group. The x-axis represents age in years at randomization, ranging from 20 to 80. The y-axis represents percent. The histogram is overlaid with a normal distribution curve, and the bars represent the distribution of ages by treatment group for "CAPECITABINE" and "S-FU + LEUCOVORIN."
Figure 3: Kaplan-Meier overall survival curves by treatment arm (datacut i6600d)

Figure 4: Empirical estimate of the hazard function for time to death events (datacut i6600d)

Figure 5: Cumulative hazard function by treatment arm (datacut i6600d)

Figure 6: Log-cumulative hazard function by treatment arm (datacut i6600d)
Standard parametric survival outcomes: The parametric functions used to plot survival according to the initial data cut are displayed in Table 2. The resulting survival curves are displayed in Figures 7-16. Figures 7-11 display the parametric hazard function that should be consistent in shape with the estimated hazard function displayed in Figure 4. When compared to figure 4 the weibull, log-logistic, and gamma appear to show similar pattern of increasing rate of the events, then a decrease in the rate of increase. The same set of parametric functions also appear to show reasonable fit when parametric survival is compared with the Kaplan-Meier curve (Figures 12-16). The visual fit is evident before approximately 2 years when there is heavy censoring. When the criteria of fit include AIC scores (Table 3), the log-logistic function has the lowest value (AIC: 2607.58).

Table 2: Parametric survival equations for each treatment arm based on first data cut

<table>
<thead>
<tr>
<th>Parametric survival model</th>
<th>Parameterized functions</th>
</tr>
</thead>
<tbody>
<tr>
<td>5FU+Leucovorin</td>
<td>Capecitabine</td>
</tr>
<tr>
<td>Exponential</td>
<td>( \exp[- \exp(-[0.0783]) \times t_i] )</td>
</tr>
<tr>
<td>Weibull</td>
<td>( \exp[- \exp\left( - \frac{0.0469}{1.4357} \times t_i^{1.4357} \right] )</td>
</tr>
<tr>
<td>Log-normal</td>
<td>( 1 - \varphi \times \left[ \frac{\log(t_i) - (2.1955)}{1.3373} \right] )</td>
</tr>
<tr>
<td>Log-logistic</td>
<td>( \frac{1}{1 + \exp\left( -[0.0474] \times t_i^{1.5480} \right]} )</td>
</tr>
<tr>
<td>Gamma</td>
<td>( \frac{1}{0.9464^2} \times \left( \exp\left( -(0.2196) \right)^{1.6733} \right) )</td>
</tr>
</tbody>
</table>
Figure 7: XACT Exponential parametric hazard function estimate (datacut i6600d)

Figure 8: XACT Weibull parametric hazard function estimate (datacut i6600d)

Figure 9: XACT Logistic parametric hazard function estimate (datacut i6600d)

Figure 10: XACT Lognormal parametric hazard function estimate (datacut i6600d)

Figure 11: XACT Gamma parametric hazard function estimate (datacut i6600d)
Figure 12: Exponential survival function by treatment arm (datacut i6600d)

Figure 13: Weibull survival function by treatment arm (datacut i6600d)

Figure 14: Log-logistic survival function by treatment arm (datacut i6600d)

Figure 15: Log-normal survival function by treatment arm (datacut i6600d)

Figure 16: Gamma survival function by treatment arm (datacut i6600d)
Table 3: AIC scores for modelled parametric survival curves for XACT trial participants

<table>
<thead>
<tr>
<th>Parametric function</th>
<th>AIC score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exponential</td>
<td>2666.00</td>
</tr>
<tr>
<td>Weibull</td>
<td>2610.71</td>
</tr>
<tr>
<td>Log-logistic</td>
<td>2607.58</td>
</tr>
<tr>
<td>Log-normal</td>
<td>2630.97</td>
</tr>
<tr>
<td>Gamma</td>
<td>2611.57</td>
</tr>
</tbody>
</table>

*Mixture-cure model*

The SEER dataset used to estimate the cure-fraction had 4,302 registered stage III colorectal cancer cases diagnosed between 1998-2001. Of these 131 (3%) had a survival time recorded as ‘0’. After excluding these, there remained 4,171 cases. Between January 1998 to December 2013, 72% of cases had a recorded event. The median survival in this population was 5 years (95% CI 4.75 – 5.25 years), with 25% of events occurring at 2 years (95% CI 1.83 – 2.08 years) (Figure 17). The parametric functions used to estimate survival in this population are displayed in figures 18 – 22. Of these functions the gamma, log-normal and log-logistic survival curves appear to remain within the bounds of the 95% confidence limits of the Kaplan-Meier curve. Similarly, the parametric hazards for the log-normal and the log-logistic functions (Figures 27 and 28) appear to show a similar shape as the estimate observed hazard (Figures 23). Adding the additional criteria of AIC score (Table 4), the log-logistic parameter function has the lowest value out of all the models (AIC 13596.31).
Figure 17: Kaplan-Meier curve for SEER stage III colorectal cancer cases diagnosed between 1998-2001

Figure 18: Kaplan-Meier vs Weibull parametric survival model for SEER stage III colorectal cancer

Figure 19: Kaplan-Meier vs Exponential parametric survival model for SEER stage III colorectal cancer

Figure 20: Kaplan-Meier vs Gamma parametric survival model for SEER stage III colorectal cancer

Figure 21: Kaplan-Meier vs Log-logistic parametric survival model for SEER stage III colorectal cancer

Figure 22: Kaplan-Meier vs Log-normal parametric survival model for SEER stage III colorectal cancer
Figure 23: Estimated empirical hazard rates for SEER patients diagnosed with stage III colorectal cancer (CC) between 1998-2001

Figure 24: SEER Weibull parametric hazard function for patients diagnosed with stage III CC

Figure 25: SEER Exponential parametric hazard function for patients diagnosed with stage III CC

Figure 26: SEER Gamma parametric hazard function for patients diagnosed with stage III CC

Figure 27: SEER Log-logistic parametric hazard function for patients diagnosed with stage III CC

Figure 28: SEER Log-normal parametric hazard function for patients diagnosed with stage III CC
Table 4: AIC scores for modelled parametric survival curves for SEER registered colorectal cancer stage III cases diagnosed between 1998-2001

<table>
<thead>
<tr>
<th>Parametric function</th>
<th>AIC score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exponential</td>
<td>13961.85</td>
</tr>
<tr>
<td>Weibull</td>
<td>13750.09</td>
</tr>
<tr>
<td>Log-logistic</td>
<td>13596.31</td>
</tr>
<tr>
<td>Log-normal</td>
<td>13612.20</td>
</tr>
<tr>
<td>Gamma</td>
<td>13603.80</td>
</tr>
</tbody>
</table>

Parameter estimates for the mixture-cure models: The set of mixture-cure equations used to plot overall survival for each treatment arm are displayed in Table 5. For the background mortality term $S^b(t)_j$ in each equation, survival for each clinical trial participant in the absence of colorectal cancer stage III is displayed in Figure 1 (blackline). The first mixture-cure model was used to estimate the cure fraction $\pi$ based on the SEER population (Table 6). Based on the fit of the SEER population to the parametric survival and hazards curves, as well as the AIC criteria, the vector of parameters based on the log-logistic model was used to seed the SEER log-likelihood mixture-cure model $\log L_{SEER}(\theta, \pi|t, \delta)$. Maximum likelihood estimate for the cure fraction was 0.58 (Table 6).

The second set of mixture-cure models was used to re-estimate the log-logistic vector of parameters, $\theta^*$, in each treatment arm based on an initial cure fraction, $\pi$, estimate of 0.58 (Table 6). The maximized log-likelihood mixture cure model equation for 5FU+Leucovorin was $\log L_{5FU+Lev}(\theta^*|t, \delta, \pi)$. The initial set of parameters, $\delta$, were 1.9461 and 0.6286. For the Capecitabine arm the log-likelihood maximized equation was $\log L_{Cap}(\theta^*|t, \delta, \pi)$ with an initial set of parameters, $\delta$, of 2.0330 and 0.6468. The re-estimated log-logistic
parameters, $\theta^*$, for a 0.58 cure fraction estimate was 3.1642 and 0.8889 for the 5FU+leucovorin treatment arm, and 3.462 and 0.9614 for the Capecitabine arm (Table 6). As the cure fraction was re-calibrated upward for each treatment arm to fit the survival curve based on the initial data cut, the parameter values decreased to 2.5422 and 0.6755 for the 5FU+Leucovorin arm, under the assumption of a cure fraction of 0.69. For the Capecitabine arm the re-estimated values were 2.3769 and 0.6128 for an assumed cure fraction of 0.78.

Table 5: Mixture-cure model parametric overall survival equations

<table>
<thead>
<tr>
<th>CF</th>
<th>5FU+Leucovorin</th>
<th>Capecitabine</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.58</td>
<td>$S_{5FU+Lev}(t_i) = S^\theta(t)_j \times \left{ \begin{array}{c} 0.58 + (1 - 0.58) \times \left( \frac{1}{1 + \exp \left( \frac{-3.1642}{0.8889} \right) t_i} \right) \ \end{array} \right}$</td>
<td>$S_{5FU+Lev}(t_i) = S^\theta(t)_j \times \left{ \begin{array}{c} 0.58 + (1 - 0.58) \times \left( \frac{1}{1 + \exp \left( \frac{-3.462}{0.9614} \right) t_i} \right) \ \end{array} \right}$</td>
</tr>
<tr>
<td>0.62</td>
<td>$S_{5FU+Lev}(t_i) = S^\theta(t)_j \times \left{ \begin{array}{c} 0.62 + (1 - 0.62) \times \left( \frac{1}{1 + \exp \left( \frac{-2.9058}{0.8048} \right) t_i} \right) \ \end{array} \right}$</td>
<td>$S_{5FU+Lev}(t_i) = S^\theta(t)_j \times \left{ \begin{array}{c} 0.62 + (1 - 0.62) \times \left( \frac{1}{1 + \exp \left( \frac{-2.7309}{0.7440} \right) t_i} \right) \ \end{array} \right}$</td>
</tr>
<tr>
<td>0.65</td>
<td>$S_{5FU+Lev}(t_i) = S^\theta(t)_j \times \left{ \begin{array}{c} 0.65 + (1 - 0.65) \times \left( \frac{1}{1 + \exp \left( \frac{-2.7309}{0.7440} \right) t_i} \right) \ \end{array} \right}$</td>
<td>$S_{5FU+Lev}(t_i) = S^\theta(t)_j \times \left{ \begin{array}{c} 0.65 + (1 - 0.65) \times \left( \frac{1}{1 + \exp \left( \frac{-2.5422}{0.6755} \right) t_i} \right) \ \end{array} \right}$</td>
</tr>
<tr>
<td>0.69</td>
<td>$S_{5FU+Lev}(t_i) = S^\theta(t)_j \times \left{ \begin{array}{c} 0.69 + (1 - 0.69) \times \left( \frac{1}{1 + \exp \left( \frac{-2.5422}{0.6755} \right) t_i} \right) \ \end{array} \right}$</td>
<td>$S_{5FU+Lev}(t_i) = S^\theta(t)_j \times \left{ \begin{array}{c} 0.69 + (1 - 0.69) \times \left( \frac{1}{1 + \exp \left( \frac{-2.7533}{0.8048} \right) t_i} \right) \ \end{array} \right}$</td>
</tr>
<tr>
<td>0.75</td>
<td>$S_{5FU+Lev}(t_i) = S^\theta(t)_j \times \left{ \begin{array}{c} 0.75 + (1 - 0.75) \times \left( \frac{1}{1 + \exp \left( \frac{-2.6022}{0.7440} \right) t_i} \right) \ \end{array} \right}$</td>
<td>$S_{5FU+Lev}(t_i) = S^\theta(t)_j \times \left{ \begin{array}{c} 0.75 + (1 - 0.75) \times \left( \frac{1}{1 + \exp \left( \frac{-2.3769}{0.6755} \right) t_i} \right) \ \end{array} \right}$</td>
</tr>
<tr>
<td>0.78</td>
<td>$S_{5FU+Lev}(t_i) = S^\theta(t)_j \times \left{ \begin{array}{c} 0.78 + (1 - 0.78) \times \left( \frac{1}{1 + \exp \left( \frac{-2.3769}{0.6755} \right) t_i} \right) \ \end{array} \right}$</td>
<td>$S_{5FU+Lev}(t_i) = S^\theta(t)_j \times \left{ \begin{array}{c} 0.78 + (1 - 0.78) \times \left( \frac{1}{1 + \exp \left( \frac{-2.3769}{0.6755} \right) t_i} \right) \ \end{array} \right}$</td>
</tr>
<tr>
<td>Population</td>
<td>Parametric Model</td>
<td>Seed parameters</td>
</tr>
<tr>
<td>---------------------</td>
<td>--------------------------------</td>
<td>-----------------</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\lambda$</td>
</tr>
<tr>
<td>SEER Stage III</td>
<td>Log-logistic</td>
<td>1.6762</td>
</tr>
<tr>
<td></td>
<td>Log-logistic (scenario 1)</td>
<td>1.9461</td>
</tr>
<tr>
<td></td>
<td>Log-logistic (scenario 2)</td>
<td>1.9461</td>
</tr>
<tr>
<td></td>
<td>Log-logistic (scenario 3)</td>
<td>1.9461</td>
</tr>
<tr>
<td></td>
<td>Log-logistic (scenario 4)</td>
<td>1.9461</td>
</tr>
<tr>
<td>5FU+Leucovorin</td>
<td>Log-logistic</td>
<td>2.0330</td>
</tr>
<tr>
<td></td>
<td>Log-logistic (scenario 2)</td>
<td>2.0330</td>
</tr>
<tr>
<td></td>
<td>Log-logistic (scenario 3)</td>
<td>2.0330</td>
</tr>
<tr>
<td>Capecitabine</td>
<td>Log-logistic</td>
<td>2.0330</td>
</tr>
<tr>
<td></td>
<td>Log-logistic (scenario 2)</td>
<td>2.0330</td>
</tr>
<tr>
<td></td>
<td>Log-logistic (scenario 3)</td>
<td>2.0330</td>
</tr>
<tr>
<td></td>
<td>Log-logistic (scenario 4)</td>
<td>2.0330</td>
</tr>
</tbody>
</table>

* The cure fraction is a fixed estimate in the maximization of the log-likelihood; $\pi$ cure-fraction; $\lambda$ is the intercept term ($\beta_0$) for the 5FU+Leucovorin arm, or the intercept plus coefficient value for the Capecitabine arm ($\beta_0 + \beta_1x$); $p$ is the scale parameter.
Assessing the best-fit parametric model

Figures 29-32 display the Kaplan-Meier curves for the initial data cut for each treatment arm, as well as the fitted parametric and mixture-cure survival curves. For the fitted curves (classic and mixture-cure) survival were extrapolated to 10 years. Among the standard set of parametric curves, the log-logistic model displays the best fit to the initial data cut of observed survival (Figures 29 and 30) when using the visual comparison between observed versus modelled functions (survival and hazard) and AIC criteria. The best fitting mixture-cure model for the 5FU+Leucovorin arm was based on a cure-fraction estimate of 0.69 (Figure 31), while for the Capecitabine arm the best fit model was based on a cure-fraction estimate of 0.78 (Figure 32). Given the intent of the mixture-cure model is to account for the potential cured proportion, which is consistent for patients diagnosed with colorectal cancer treated with curative intent, the mixture-cure models were selected as the best predictor of survival up to 10 years. Among the modelled mixture-cure models, the cure-fraction of 0.69 is my prediction of the best fit for the 5FU+Leucovorin arm, while a cure-fraction of 0.78 is my prediction of the best fit for the Capecitabine arm.
Figure 29: 5FU+Leucovorin treatment arm comparison of non-parametric and extrapolated parametric survival curves.
Figure 30: Capecitabine treatment arm comparison of non-parametric and extrapolated parametric survival curves
Figure 31: 5FU+Leucovorin treatment arm comparison of non-parametric and extrapolated mixture cure parametric survival curves by cure fraction.
Figure 32: Capecitabine treatment arm comparison of non-parametric and extrapolated mixture cure parametric survival curves by cure fraction.
Model validation

Non-parametric survival outcomes of final XACT trial data-cut: Figures 33-35 display the non-parametric survival outcomes for each treatment arm based on the last datacut (i66001h). Median survival was not reached during the final datacut (Figure 33). At the final datacut, 25% of the sample had an event in the 5FU+Leucovorin arm at 3.70 years (95% CI 3.16 – 4.34 years), while for the Capecitabine arm the same proportion of events occurred at 4.21 yrs (95% CI 3.71 – 4.80 ). While there appears to be a difference between the treatment arms in the time point at which 25% of the sample experienced an event, this difference was not statistically significant (Log-rank 3.54, p-value 0.0598; Wilcoxon test value 2.76, p-value 0.0961).

The difference in the hazard rate at which events occurred in each treatment arm appears to show an increasing rate of death until approximately 2.5-3 years where the peak is reached (Figure 34). After this point, the rate of events declines at a similar rate in each treatment arm. However in both arms the rates tend to pick-up in the tail of the curves, yet these may be unreliable given the small number of events at this point. Figure 35 also provides insight into the rate at which the hazard is occurring. The log of the cumulative hazard relative to the log of time shows a slower rate of events initially, then an increase in the rate subsequently with a slight divergence at 2.5 yrs (ln(0.92) on the x-axis of figure 35) between the two treatment arms.

Best-fit parametric survival curves vs observed survival: Figures 36 and 37 display the observed survival for the final datacut for each treatment arm and the corresponding best fit parametric model based on the classic parametric and the mixture-cure models. The classic parametric models are predictions of the final datacut based on the observed events in the first datacut. The mixture-cure model with a cure-fraction of 0.69 appears to be a better fit compared with the observed survival in the 5FU+Leucovorin treatment arm (Figure 33). In the initial stages
of observation, the log-logistic curve appears to follow the observed survival at approximately 3.5 years. However, as more patients are observed, the log-logistic prediction becomes too aggressive relative to the observed survival. In contrast the mixture-cure fraction while initially over-estimating the failure rate, appears to converge towards a similar trajectory as the observed survival.

For the Capecitabine arm (Figure 34), the mixture-cure model with a cure fraction of 0.78 appears to be the best fit compared with the observed overall survival. Similar to the initial trajectory for the 5FU+Leucovorin arm, the log-logistic model appears to display a better fit upto observed survival at approximately 4 years. However, the trajectory becomes too aggressive and subsequently overestimates the failure proportion as time progresses. In contrast, the mixture-cure model follows the trajectory of the observed survival proportion, justifying its choice as a better parametric fit.
Figure 33: Kaplan-Meier overall survival curves by treatment arm (datacut i6600d)

Figure 34: Empirical estimate of the hazard function for time to death events (datacut i6600d)

Figure 35: Log-cumulative hazard function by treatment arm (datacut i6600d)
Figure 36: 5FU+Leucovorin treatment arm final data cut observed vs predicted overall survival

- Survival probability vs Time (Years)
- Loglogistic model
- CF = 0.69
Figure 37: Capecitabine treatment arm final data cut observed vs predicted overall survival

Survival probability vs time (years) for Capecitabine KM final data cut and two different models: loglogistic and a model with CF=0.78.
**Extrapolation beyond observed survival:** Beyond the period of observed survival time, extrapolated survival for the log-logistic model follows a different survival trajectory from that of the extrapolated mixture-cure model. For the 5FU+Leucovorin arm (purple line – Figure 38) the observed survival for the initial data-cut is followed closely using the predicted log-logistic survival function (dashed orange line). This is the same for the Capecitabine arm (green line – Figure 39). The predicted log-logistic survival trajectory diverges from observed survival for the final data-cut (5FU+Leucovorin - red line, Figure 38; Capecitabine – black line, Figure 39), illustrating a more aggressive probability of death. In contrast, the re-calibrated mixture-cure models initially underestimate the observed survival for both treatment arms, but then begins to follow the trajectory for the final data-cut. As the trajectory of the log-logistic model continues the survival probability tail at older years (between 30 – 50 years after age at randomization) displays overly optimistic outcome, such that the probability of death occurs at a slower rate at older ages. In contrast the mixture-cure model displays a probability survival that is less aggressive due to the cured proportion who have a trajectory that is consistent with their age that is similar to the background population. Towards the tail the mixture-cure model is less optimistic with a small proportion of patients alive at 50 years after age at randomization. This is likely to reflect the small proportion of patients who are less than 55 years of age.
Figure 38: 5FU+Leucovorin treatment predicted modelled long term survival vs observed survival
Figure 39: Capecitabine 5FU+Leucovorin treatment predicted modelled long term survival vs observed survival
Discussion

My objectives in this thesis were to firstly fit a set of parametric and mixture-cure models to overall survival in patients diagnosed with stage III colorectal cancer treated with curative intent, based on an initial data-cut of the XACT clinical trial. Secondly, I assessed which set of parametric models (classic and mixture-cure) best fit the observed overall survival based on the same initial data-cut. Finally, I evaluated the selected best fit parametric models predicted observed overall survival based on the trials final data-cut, and used these to derive long-term survival estimates. My main findings in this analysis were that classic parametric models are unlikely to accurately reflect observed overall survival in the XACT clinical patients diagnosed with stage III colorectal cancer treated in the adjuvant setting. The best fit parametric function (log-logistic) appeared to fit well to the overall survival based on the initial XACT trial data-cut. However, when compared with the trial’s final data-cut, the predicted log-logistic overall survival trajectory overestimated stage III colorectal cancer failure probability. In contrast, when the mixture cure model was calibrated to the initial XACT trial data-cut, its predicted overall survival trajectory appeared to broadly follow the pattern of observed overall survival calculated using the final data-cut.

A key component of this analysis was the use of the mixture-cure model and the derivation of the cure fraction estimate. For this analysis I used the SEER population of patients diagnosed with stage III colorectal cancer between 1998-2001. The initial estimated cure fraction, when used in the mixture-cure model to derive predicted survival, dramatically underestimated the observed survival of the clinical trial population based on the initial data-cut. This resulted in the need to recalibrate the cure-fraction so the predicted survival approximated the observed survival (of the initial data-cut). The underestimate of the cure-fraction based on the SEER
population presents one of the issues of using an external population which may not be similar to the clinical trial population in which a cost-effectiveness model is being developed. The difference was evident when non-parametric survival curves for the clinical trial population had a higher survival probability at the 4-5 year follow-up period (60-80%) compared with the SEER stage III population (50-60%). To ensure that an external population used to derive the cure-fraction is broadly similar with the clinical trial population to be modelled, an exploratory analysis using Cox proportional hazards modeling could be used to assess which variable may explain the difference between the two populations.

Extrapolation of survival was the final part of the analysis for my thesis. Predicted survival was modelled for 50 years at age of randomization for each treatment arm using the best-fit parametric model and the re-calibrated mixture cure model. In this thesis I argue that the mixture-cure model is the best-fit model when the final datacut for observed survival is used. In contrast, the parametric model selected as the best fit, while initially the better fit based on the first datacut, over-estimates the probability of failure. Consequently, when the mixture-cure and the log-logistic survival models are projected beyond the observed survival, the two sets of trajectories provide striking different survival patterns. The log-logistic model overestimates the probability of failure in the initial period after randomization between 0-20 years, however, towards the 30-50 year period predicted survival is overly optimistic. The long tails of the log-logistic and log-normal models is one of their limitations in extrapolating predicted survival. In contrast the mixture-cure model takes into account a proportion of the population who are cured. The survival trajectory therefore captures outcomes that reflect background mortality for certain proportion of the clinical trial population. This results in a predicted survival trajectory that is more optimistic in the initial period after randomization, which then becomes more aggressive towards the tail. However,
it should be borne in mind that the survival trajectory for a mixture-cure model will be
dependent on the age characteristics of the trial population. A different age population
structure will yield different long-term mixture-cure predicted survival.

The final area of discussion for the thesis I would like to address is around ‘uncertainty’.
Sampling variability was assessed in this thesis through simple approaches such as the 95%
confidence interval for the observed Kaplan-Meier curves. However, for each of the
parametric and mixture-cure models, point estimate of the parameters were used to estimate
predicted survival. One of the limitations of this thesis is that the uncertainty around these
point estimates (alone and combined) were not explored in great detail. The method of
maximum likelihood using PROC NLMIXED provides standard errors for each of the
parameters. To have addressed the uncertainty further I would have like to have used monte-
carlo simulation to explore the combined uncertainty with each set of parameter estimates in
the parametric and mixture-cure models. Using the range of point estimates from the monte
carlo simulation, I would apply each of these to the time points for each predicted event.
While I believe the direction of the predicted survival trajectory would not have changed, the
range of the estimates of the projected survival trajectory would be worth assess for the
purposes of estimating the variability in the ICER.

In addition to addressing statistical uncertainty, I would like to have compared the predicted
survival estimates with one or more population-based and/or clinical cancer registries. The
population-based registries could provide insight as to how well those diagnosed with stage
III colorectal cancer in the general population follow the predicted long term survival
trajectory based on the XACT clinical trial patient characteristics. If there were differences
(and there are likely to be differences based on the SEER and XACT trial Kaplan-Meier
curves), multivariate analysis using Cox regression could be used to assess which clinical characteristics (covariates) may/may not explain the difference between the predicted and actual survival trajectories.
Appendix

/***********************************************
/*MACRO to generate standard parametric survival curves*/
/***********************************************

/*Macro for for the 1st datacut starts here i66001d.pbe*/

/***********************************************
** DESCRIPTION: Assessment of model fit to 1st data cut - Xeloda
**
** INPUTS: work.effic4
**
** OUTPUTS: Range of STEM outputs
**
** NOTES: This work is completed for a Masters Biostats thesis
**
***********************************************

** LOCATION : $Source: / $
**
** VERSION : $Id: $
**
** AUDIT :

***********************************************

/** Initialise macros/set options and libnames

***********************************************

/*sets general option and includes all MORSE macros*/
%include "$PROD/morse/sas/morse_init.sas";

%let pop_in=ITT; *Population used in the analysis;*/
%let treat_us=rnd; *Treatment used in the analysis;*/
%let cut_off=11 Feb 2014;*/

proc datasets lib = work mt = data kill nowarn nolist;
run;
quit;

/** Arrange data to be read in stem macro**/

/** This is the efficacy data ***/;

data event_b;
set ana.effic;
  label rndshortn = "Randomized Treatment Group"
  ttdfs_m = "Time to disease relapse (months)"
  ttdied_m = "Time to death (months)"
  ttdied_m = ttdied / (365.25/12);
  ttdfs_m = ttdfs / (365.25/12);
  ttdfs_yrs=ttdfs_m/12;
  ttdied_yrs = ttdied/356.25;
  pop_itt=1;
  if rnd = "CAPECITABINE" then rndshortn = 2;
  else if rnd = "5-FU + LEUCOVORIN" then rndshortn = 1;
run;
proc sort data = work.event_b;
   by proto crtn pt;
run;
/**
variable name definitions ana.effic:
proto = protocol
crtn = clinical research task number
pt = patient number
**/ 

/*** This is the demographic data ***/;
data demo_b;
   set ana.demoext;
   keep proto crtn pt age sex trt1 trt1dt rnd s_cntry rndshortn pop_itt;
   pop_itt=1;
   if rnd = "CAPECITABINE" then rndshortn = 2;
   else if rnd = "$5-FU + LEUCOVORIN" then rndshortn = 1;
run;
proc sort data = work.demo_b;
   by proto crtn pt;
run;
/**
variable name definitions for ana.demoext:
proto = protocol
crtn = clinical research task number
pt = patient number
trt1 = first trial medication
trt1dt = first trial medicatoin begin SAS datetime
rnd = randomisation treatment
**/ 
run;

*** create a new document ***;
options noquotelenmax;
%util_document_new(document_name = stem); run;

%stem_macro(dsin          = event_b
demo          = demo_b
dsini where    = pop_itt=1
demomega where = pop_itt=1
usubjid       = pt
rxc           = rnd
rxc           = rnd
stratvar      =
timevar       = ttdied_yrs
censvar       = censdied
censval       = 0
document_name = stem
outdoc_prefix = os
title1        = "Study: XACT, Population: Adjuvant treatment stage III colon cancer"
title2        = "Duration of overall Survival"
footnotel     = "Cap = CAPECITABINE"
footnote2     = "$5fu_leu = 5-FU + LEUCOVORIN"
);
run;

/*** output the selected outputs ***/;
%util_document_play(document_name = stem

outpath       = $HOME/cd10743d.pbe/i66001d.pbe/reports
exl_all_os.pdf
[
  os_km_pl
  os_lr_inf
  os_lrkm_pl_est
  os_lr_est
  os_lr_est_expo
  os_lr_est_weib
  os_lr_est_llog
  os_lr_est_gamm
  os_lr_est_lnor
  os_km_plt_h
  os_km_plt_s
  os_km_plt_ls
  os_km_plt_lls
  os_lrkm_plt_expo
  os_lrkm_plt_weib
  os_lrkm_plt_llog
  os_lrkm_plt_lnor
  os_lrkm_plt_gamm
  os_lr_plt_expo
  os_lr_plt_weib
  os_lr_plt_llog
  os_lr_plt_lnor
  os_lr_plt_gamm
  os_lrhaz_plt_expo
  os_lrhaz_plt_weib
  os_lrhaz_plt_llog
  os_lrhaz_plt_lnor
  os_lrhaz_plt_gamm
  os_lrhaz_plt_gomp
  os_lrhaz_plt_nphw
]
exl_raw_os.pdf
[
  os_km_raw
  os_lr_raw_expo
  os_lr_raw_weib
  os_lr_raw_lnor
  os_lr_raw_gamm
  os_lr_raw_llog
]
exl_all_os.xls
[
  os_km_pl
  os_lr_inf
  os_lrkm_pl_est
  os_lr_est
  os_lr_est_expo
  os_lr_est_weib
  os_lr_est_llog
  os_lr_est_gamm
  os_lr_est_lnor
]
exl_raw_dfs.xls
[
  os_km_raw
  os_lr_raw_expo
  os_lr_raw_weib
  os_lr_raw_lnor
  os_lr_raw_gamm
  os_lr_raw_llog
]

);
quit;
%mend stem_macro;
References


