

# Introduction

*Philip Cooney*

*16 September 2019*

## Contents

<b>1</b>	<b>Chapter Overview</b>	<b>2</b>
<b>2</b>	<b>Overview of Survival Analysis</b>	<b>2</b>
2.1	Features of survival data . . . . .	2
2.2	Practical Considerations in survival analysis . . . . .	3
2.3	Observed Survivor function . . . . .	4
2.4	Censoring and Truncation mechanisms . . . . .	5
<b>3</b>	<b>Fundamental functions in Survival Analysis</b>	<b>6</b>
3.1	Survival function . . . . .	6
3.2	Hazard and Cumulative hazard function . . . . .	7
3.3	Likelihood of right censored data . . . . .	8
<b>4</b>	<b>Non-Parametric analysis of Survival data</b>	<b>9</b>
<b>5</b>	<b>Semi-Parametric models</b>	<b>10</b>
<b>6</b>	<b>Parametric models</b>	<b>11</b>
6.1	Exponential Distribution . . . . .	11
6.2	Weibull Distribution . . . . .	12
6.3	Gompertz Distribution . . . . .	13
6.4	Log-Logistic Distribution . . . . .	15
<b>7</b>	<b>Assessing the suitability of a parameteric model</b>	<b>17</b>
7.1	Weibull or Exponential . . . . .	18
7.2	Gompertz . . . . .	19
7.3	Log-logistic . . . . .	21
<b>8</b>	<b>Overview of Health Economics</b>	<b>22</b>
8.1	Cost-effectiveness analysis (CEA) . . . . .	22
8.2	Cost-benefit analysis (CBA) . . . . .	22
8.3	Cost-effectiveness plane . . . . .	23
8.4	Valing Health - QALY . . . . .	24

8.5	Cost effectiveness models . . . . .	24
8.6	Expected Health Benefits and Costs . . . . .	25
<b>References</b>		<b>25</b>

## 1 Chapter Overview

In this opening chapter I will provide an introduction to survival analysis, along with a description of the characteristics of survival data and some of the statistical methods which can be used to analyze such data. Additionally because this thesis is researching techniques to extrapolate time to event outcomes for the purposes of health technology assessment, a brief overview of health economic will be presented.

## 2 Overview of Survival Analysis

The primary aim of many studies is to analyze the time until a prespecified event of interest occurs. In these settings, the response variable is the time until that event, which is often called failure time, survival time, or event time.

Survival analysis is most heavily used in clinical and epidemiologic studies, in which the event may be death, the appearance of a tumor, the development of some disease, recurrence of a disease, conception, or cessation of smoking.

However, event times also are frequently encountered in other disciplines. For example, in sociology we could be interested in the duration of a first marriage, in marketing the length of subscription to a newspaper or a magazine, and in industry the time a component of machine operates without any failures. Although in these situations the term “Time to Event” is more general, the term Survival analysis will be used throughout this document.

Although the analysis of time to event is broadly characterised as Survival analysis, the statisitcal techniques employed are typically context specific. In clinical trials which have a time to event outcome, the primary objective is to identify if there is a statisitcal significant difference in the expected survival times of the treatment arm. In other disciplines such as health economics the focus is to assess the long term expected survival of a treatment group. Other researchers may be interested in how several longitudinally measured responses and the time until an event interest occurs are related. In another context a researcher may be interested in the outcomes of related individuals, such as twins with respect to some variable of interest.

In this PhD, the focus is the assessment of expected survival beyond that observed in a clinical trial, however, I (hope) to employ some of the statistical methods from other areas of survival analysis to facilitate the prediction of long term survival.

### 2.1 Features of survival data

Before introducing notation, I will first consider the reasons why survival data are not amenable to standard statistical procedures used in data analysis. One reason is that survival data are generally not symmetrically distributed. Typically, a histogram constructed from the survival times of a group of similar individuals will tend to be positively skewed, that is, the histogram will have a longer tail to the right of the interval that contains the largest number of observations. Secondly the time to event data are typically not observed for all observations under study, and the time to events for these observations are *censored*. Typically this occurs because the data from a study are to be analysed at a point in time when some individuals are still alive. Alternatively, the survival status of an individual at the time of the analysis might not be known because that individual has been lost to *follow-up*.

## 2.2 Practical Considerations in survival analysis

### 2.2.1 Study time and patient time

In a typical study, patients are not all recruited at exactly the same time, but accrue over a period of months or even years. After recruitment, patients are followed up until they die, or until a point in calendar time that marks the end of the study, when the data are analysed. Although the actual survival times will be observed for a number of patients, after recruitment some patients may be lost to follow-up, while others will still be alive at the end of the study.

The calendar time period in which an individual is in the study is known as the study time. The study time for eight individuals in a clinical trial is illustrated diagrammatically in Figure 1, in which the time of entry to the study is represented by a (●).

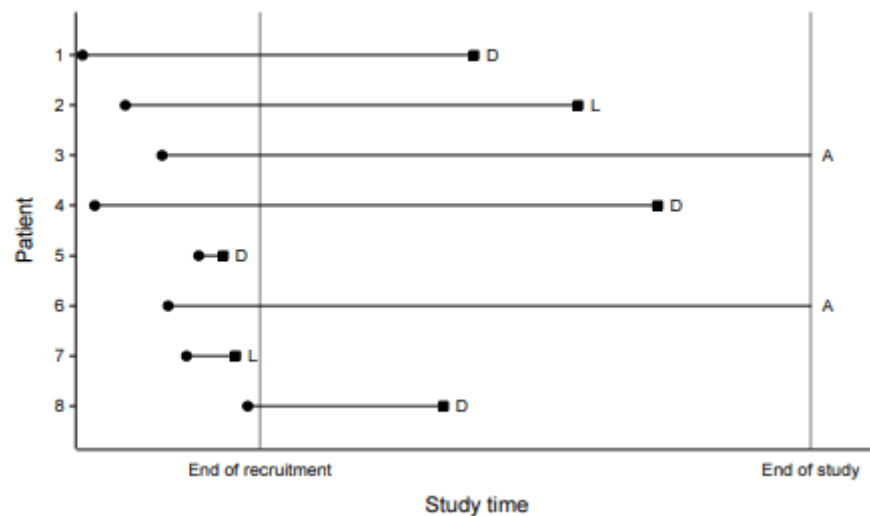


Figure 1: Study time for eight patients in a survival study

This figure shows that individuals 1, 4, 5 and 8 die (D) during the course of the study, individuals 2 and 7 are lost to follow-up (L), and individuals 3 and 6 are still alive (A) at the end of the observation period. As far as each patient is concerned, the trial begins at some time  $t_0$ . The corresponding survival times for the eight individuals depicted in Figure 1 are shown in order in Figure 2. The period of time that a patient spends in the study, measured from that patient's time origin, is often referred to as patient time. The period of time from the time origin to the death of a patient (D) is then the survival time, and this is recorded for individuals 1, 4, 5 and 8. The survival times of the remaining individuals are right-censored (C).

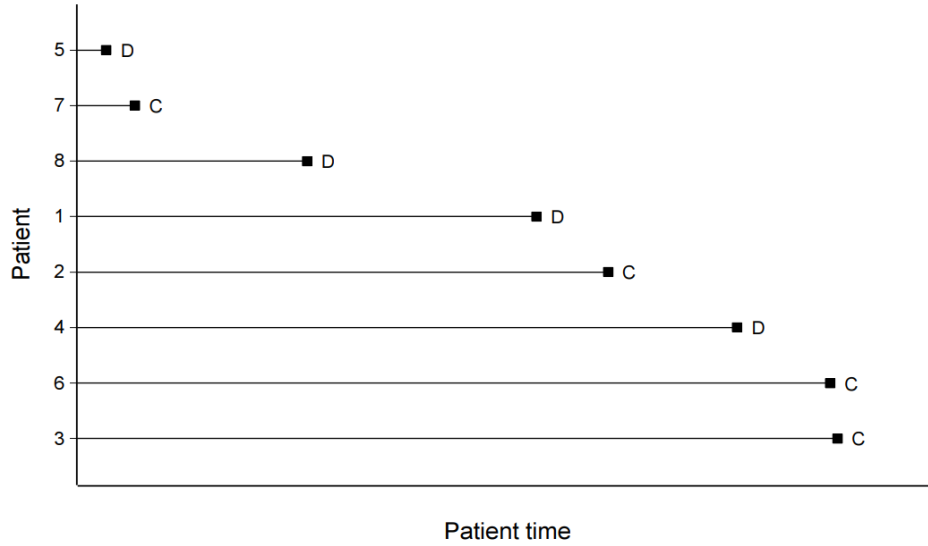


Figure 2: Patient time for eight patients in a survival study

In practice, the actual data recorded will be the date on which each individual enters the study, and the date on which each individual dies or was last known to be alive. The survival time in days, weeks or months, whichever is the most appropriate, can then be calculated. Most computer software packages for survival analysis have facilities for performing this calculation from input data in the form of dates

## 2.3 Observed Survivor function

When using statistical models to describe survival data we usually consider the probability of survival vs time to be a smooth continuous function (to be discussed in Section 3.1). In practice, when using actual data, we usually obtain graphs that are step functions, as illustrated in Figure 3, rather than smooth curves. Moreover, because the study period is never infinite in length and there may be competing risks for failure, it is possible that not everyone studied gets the event. The estimated survivor function, denoted by a caret over the S in the graph, thus may not go all the way down to zero at the end of the study.

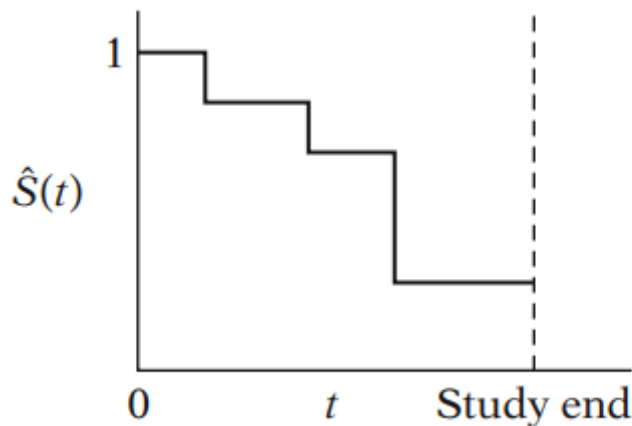


Figure 3:  $\hat{S}(t)$  in practice

## 2.4 Censoring and Truncation mechanisms

As discussed previously, censoring occurs when some lifetimes are known to have occurred only within certain intervals. The remainder of the lifetimes are known exactly. There are various categories of censoring, such as right censoring, left censoring, and interval censoring.

A second feature which may be present in some survival studies is that of *truncation*. Truncation of survival data occurs when only those individuals whose event time lies within a certain observational window  $(Y_L, Y_R)$  are observed. An individual whose event time is not in this interval is not observed and no information on this subject is available to the investigator. This is in contrast to censoring where there is at least partial information on each subject. Because we are only aware of individuals with event times in the observational window, the inference for truncated data is restricted to conditional estimation. The main impact on the analysis, when data are truncated, is that the investigator must use a conditional distribution in constructing the likelihood, or employ a statistical method which uses a selective risk set. For further details see (J. Klein and Moeschberger 2003).

### 2.4.1 Right Censoring

First, we will consider Type I censoring where the event is observed only if it occurs prior to some prespecified time. These censoring times may vary from individual to individual. A typical clinical trial starts with a fixed number of animals or patients to which a treatment (or treatments) is (are) applied. Because of time or cost considerations, the investigator will terminate the study or report the results before all subjects realize their events.

### 2.4.2 Left Censoring

A lifetime  $X$  associated with a specific individual in a study is considered to be *left censored* if it is less than a censoring time  $C_l$  ( $C_l$  for left censoring time), that is, the event of interest has already occurred for the individual before that person is observed in the study at time  $C_l$ . For such individuals, we know that they have experienced the event sometime before time  $C_l$ , but their exact event time is unknown. A possible example of this is HIV exposure, patients may not know exactly when they were exposed to the virus, but will know that it was previous to the virus becoming symptomatic.

### 2.4.3 Interval Censoring

A more general type of censoring occurs when the lifetime is only known to occur within an interval. Such interval censoring occurs when patients in a clinical trial or longitudinal study have periodic follow-up and the patient's event time is only known to fall in an interval  $(L_i, R_i]$  ( $L$  for left endpoint and  $R$  for right endpoint of the censoring interval).

In the Framingham Heart Study, the ages at which subjects first developed coronary heart disease (CHD) are usually known exactly. However, the ages of first occurrence of the subcategory angina pectoris may be known only to be between two clinical examinations, approximately two years apart (J. Klein and Moeschberger 2003). Such observations are interval-censored.

### 2.4.4 Left truncation

Left truncation occurs when subjects enter a study at a particular age (not necessarily the origin for the event of interest) and are followed from this delayed entry time until the event occurs or until the subject is censored. We only observe those individuals whose event time  $X$  exceeds the truncation time  $Y_L$ .

### 2.4.5 Right truncation

Right truncation occurs when only individuals who have experienced the event of interest are observable. That is, we observe the survival time  $X$  only when  $X \leq Y_R$ . An example of a right-truncated sample is a mortality study based on death records.

## 3 Fundamental functions in Survival Analysis

### 3.1 Survival function

Consistent with the standard notation in survival analysis, let capital  $T$  be the random variable for a person's survival time. Since  $T$  denotes time, its possible values include all nonnegative numbers; that is,  $T$  can be any number equal to or greater than zero.

Let small letter  $t$  denote any specific value of interest for the random variable capital  $T$ . For example, if we are interested in evaluating whether a person survives for more than 5 years after undergoing cancer therapy, small  $t$  equals 5; we then ask whether capital  $T$  exceeds 5.

$$T > t = 5$$

Finally, we let the Greek letter delta ( $\delta$ ) denote a (0,1) random variable indicating either failure or censorship. That is,  $\delta = 1$  for failure if the event occurs during the study period, or  $\delta = 0$  if the survival time is censored by the end of the study period. Note that if a person does not fail, that is, does not get the event during the study period, censorship is the only remaining possibility for that person's survival time. That is,  $\delta = 0$  if and only if one of the following happens: a person survives until the study ends, a person is lost to follow-up, or a person withdraws during the study period.

We assume that  $T$  has a probability density function (p.d.f.)  $f(t)$  and cumulative distribution function (c.d.f.)  $F(t) = Pr(T < t)$ , giving the probability that the event has occurred by duration  $t$ .

The survivor function  $S(t)$  gives the probability that a person survives longer than some specified time  $t$ : that is,  $S(t)$  gives the probability that the random variable  $T$  exceeds the specified time  $t$  (i.e. the complement of the c.d.f function).

$$S(t) = Pr(T \geq t) = 1 - F(t) = \int_t^{\infty} f(x) dx \quad (1)$$

Theoretically, as  $t$  ranges from 0 up to infinity, the survivor function can be graphed as a smooth curve. As illustrated by the graph, where  $t$  identifies the X-axis, all survivor functions have the following characteristics (illustrated in Figure 4):

1. they are nonincreasing; that is, they head downward as  $t$  increases;
2. at time  $t = 0$ ,  $S(t) = S(0) = 1$ ; that is, at the start of the study, since no one has gotten the event yet, the probability of surviving past time 0 is one;
3. at time  $t = \infty$ ,  $S(t) = S(\infty) = 0$ ; that is, theoretically, if the study period increased without limit, eventually nobody would survive, so the survivor curve must eventually fall to zero.

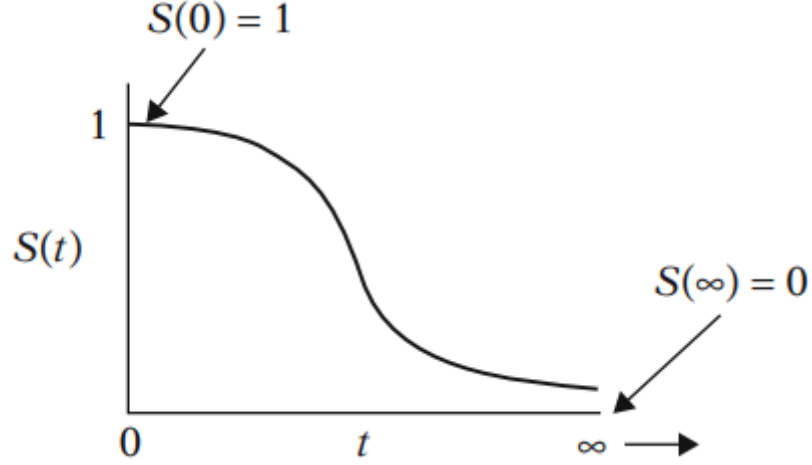


Figure 4: Theoretical properties of the survival function

Another useful statistic that can be derived from Survival function is the mean  $\mu$  or expected value of  $T$ . By definition, the expectation of a random variable is calculated by multiplying  $t$  by the density  $f(t)$  and integrating, so

$$\mu = \int_0^{\infty} t f(t) dt$$

Integrating by parts, it can be shown that

$$\mu = \int_0^{\infty} S(t) dt \quad (2)$$

### 3.2 Hazard and Cumulative hazard function

The hazard function, denoted by  $h(t)$ , is given by the formula:  $h(t)$  equals the limit, as  $t$  approaches zero, of a probability statement about survival, divided by  $t$ , where  $t$  denotes a small interval of time.

$$\lambda(t) = \lim_{dt \rightarrow 0} \frac{P(t \leq T < t + dt | T \geq t)}{dt} \quad (3)$$

The hazard function  $h(t)$  gives the instantaneous potential per unit time for the event to occur, given that the individual has survived up to time  $t$ .

The conditional probability in the numerator may be written as the ratio of the joint probability that  $T$  is in the interval  $[t, t + dt)$  and  $T \geq t$  (which is, of course, the same as the probability that  $t$  is in the interval), to the probability of the condition  $T \geq t$ . The former may be written as  $f(t)dt$  for small  $dt$ , while the latter is  $S(t)$  by definition. Dividing by  $dt$  and passing to the limit gives the useful result:

$$\lambda(t) = \frac{f(t)}{S(t)} \quad (4)$$

From the definition of  $S(t)$  (Equation (1)) it can be seen that  $-f(t)$  is the derivative of  $S(t)$ . Noting that  $\int \frac{1}{x} dx = \log|x| + C$  allows us to rewrite the Equation (4) as:

$$\lambda(t) = -\frac{d}{dt} \log S(t) \quad (5)$$

With some further manipulation of (5) the following equation for  $S(t)$  is obtained:

$$S(t) = \exp\left\{-\int_0^t \lambda(x)dx\right\} \quad (6)$$

Another function that is closely related to the hazard and survival function is the cumulative hazard  $\Lambda(t)$ , which can be considered the sum of the risks you face going from duration 0 to  $t$ .

$$\Lambda(t) = \int_0^t \lambda(x)dx = -\log S(t) \quad (7)$$

An important point to note is that throughout our analysis we assume non-informative censoring. This means that the actual survival time of an individual,  $t$ , does not depend on any mechanism that causes that individual's survival time to be censored at time  $c$ , where  $c < t$ . Statistical methods in survival analysis typically make this assumption by default, because analogous to not missing at random (MNAR) longitudinal data, little meaningful analysis can be performed with the introduction of external assumptions/information.

### 3.3 Likelihood of right censored data

In constructing a likelihood function for censored or truncated data we need to consider carefully what information each observation gives us. An observation corresponding to an exact event time provides information on the probability that the event's occurring at this time, which is approximately equal to the density function of  $X$  at this time. For a right-censored observation all we know is that the event time is larger than this time, so the information is the survival function evaluated at the on study time. As stated before a critical assumption is that the lifetimes and the censoring times are independent.

Combining this information we obtain

$$L \propto \prod_{i \in D} f(x_i) \prod_{i \in R} S(C_r) \quad (8)$$

where  $D$  is the set of death times,  $R$  the set of right-censored observations.

Data from experiments involving right censoring can be conveniently represented by pairs of random variables  $(T, \delta)$ , where  $\delta$  indicates whether the lifetime  $X$  is observed ( $\delta = 1$ ) or not ( $\delta = 0$ ), and  $T$  is equal to  $X$  if the lifetime is observed and to  $C_r$  if it is right-censored, i.e.,  $T = \min(X, C_r)$ .

Details of constructing the likelihood function for *Type I censoring* are as follows. For  $\delta = 0$ , it can be seen that

$$\begin{aligned} Pr[T, \delta = 0] &= Pr[T = C_r | \delta = 0] Pr[\delta = 0] = Pr(\delta = 0) \\ &= Pr(X > C_r) = S(C_r) \end{aligned}$$

Also, for  $\delta = 1$ ,

$$\begin{aligned} Pr[T, \delta = 1] &= Pr[T = X | \delta = 1] Pr[\delta = 1] \\ &= Pr[X = T | X \leq C_r] Pr(X \leq C_r) \\ &= \left[ \frac{f(t)}{1 - S(C_r)} \right] [1 - S(C_r)] = f(t) \end{aligned}$$



These expressions can be combined into the single expression

$$Pr(t, \delta) = [f(t)]^\delta [S(t)]^{1-\delta}$$

If we have a random sample of pairs  $(T_i, \delta_i), i = 1, \dots, n$ , the likelihood function is

$$L = \sum_{i=1}^n Pr[t_i, \delta_i] = \sum_{i=1}^n [f(t_i)]^{\delta_i} [S(t_i)]^{1-\delta_i} \quad (9)$$

which is of the same form as (8). Because we can write  $f(t_i) = h(t_i)S(t_i)$  we can write this likelihood as

$$L = \prod_{i=1}^n [h(t_i)]^{\delta_i} \exp[-H(t_i)] \quad (10)$$

From this equivalent formulation it is clear that all subjects contribute an amount to the likelihood equal to the exponent of the negative of the cumulative hazard function evaluated at their corresponding observed censored or event time  $T_i$ . Subjects who experienced the event additionally contribute an amount equal to the hazard function evaluated at  $T_i$ . Thus, censored observations contribute less information to the statistical inference than uncensored observations and highlight how our uncertainty about the appropriate survival distribution increases as more of the patient sample are censored.

## 4 Non-Parametric analysis of Survival data

In order to get an understanding of how the survival evolves over time, arguably the most informative summary is the Kaplan Meier plot of the survival function. This is a nonparametric estimator that does not make any assumptions for the underlying distribution of the failure times. To introduce this estimator, let  $t_1, \dots, t_k$  denote the unique event times in the sample at hand. Using the law of total probability, the probability of surviving any time point  $t$  can be written as the product of the conditional probabilities:

$$Pr(T^* > t) = Pr(T^* > t | T^* > t-1) \times Pr(T^* > t-1 | T^* > t-2) \times \dots$$

To estimate survival probabilities at each unique event time, we utilize the above expansion, and in the calculation of the conditional probabilities, we account for censoring by suitably adjusting the number of subjects at risk (i.e., the subjects who have not experienced the event and are not censored), which leads to the estimator:

$$\hat{S}_{KM}(t) = \prod_{i:t_i \leq t} \frac{r_i - d_i}{r_i} \quad (11)$$

where  $r_i$  denotes the number of subjects still at risk at the unique event  $t_i$ , and  $d_i$  is the number of events at  $t_i$ .

The standard error of the Kaplan-Meier estimate of the survivor function, defined to be the square root of the estimated variance of the estimate, is given by

$$se\{\hat{S}(t)\} \approx \hat{S}(t) \left\{ \sum_{j=1}^k \frac{d_j}{n_j(n_j - d_j)} \right\}^{\frac{1}{2}} \quad (12)$$

for  $t_{(k)} \leq t < t_{(k+1)}$ . This result is known as Greenwood's formula and is derived from the asymptotic normality for  $\hat{S}_{KM}(t)$  and the Delta Method.

A superior estimator can be obtained from the complementary log-log transformation of the survival function  $\log\{-\log\hat{S}(t)\}$ . This ensures that the confidence limits for  $S(t)$  will not cross the boundaries of the interval  $[0, 1]$ . The variance of  $\log\{-\log S(t)\}$  is obtained through application of the Delta method to Equation (12), which gives:

$$\text{var} [\log - \log \hat{S}(t)] \approx \frac{1}{\{\log \hat{S}(t)\}^2} \sum_{j=1}^k \frac{d_j}{n_j(n_j - d_j)} \quad (13)$$

The standard error of  $\log\{-\log S(t)\}$  is the square root of this quantity. This leads to  $100(1 - \alpha)\%$  limits of the form

$$\hat{S}(t)^{\exp[\pm z_{\alpha/2} \text{se}\{\log[-\log \hat{S}(t)]\}]}$$

where  $z_{\alpha/2}$  is the upper  $\alpha/2$ -point of the standard normal distribution.

Another non-parametric estimator is the Nelson-Aalen estimator. Details can be found in the following references (J. Klein and Moeschberger 2003).

## 5 Semi-Parametric models

Arguably the most famous survival model is the Cox-proportional hazard model.

The basic model due to (Cox 1972) is as follows:

$$h(t|Z) = h_0(t) \exp(\beta^t Z) = h_0(t) \exp\left(\sum_{k=1}^p \beta_k Z_k\right) \quad (14)$$

where  $Z$  denotes a vector of covariates related to an individuals risk, with  $\beta = (\beta_1, \dots, \beta_p)^t$  a parameter vector.

This is called a semiparametric model because a parametric form is assumed only for the covariate effect. The model is known as a proportional hazards model because, if we look at two individuals with covariate values  $Z$  and  $Z^*$ , the ratio of their hazards is

$$\frac{h(t|Z)}{h(t|Z^*)} = \frac{h_0(t) \exp[\sum_{k=1}^p \beta_k Z_k]}{h_0(t) \exp[\sum_{k=1}^p \beta_k Z_k^*]} = \exp\left[\sum_{k=1}^p \beta_k (Z_k - Z_k^*)\right] \quad (15)$$

which is a constant. Therefore the hazard rates are proportional.

Estimation of this model is based on the partial likelihood (Cox 1975) as shown below

$$L(\beta) = \prod_{i=1}^D \frac{\exp[\sum_{k=1}^p \beta_k Z_{(i)k}]}{\sum_{j \in R(t_i)} \exp[\sum_{k=1}^p \beta_k Z_{jk}]} \quad (16)$$

where  $t_1 < t_2 < \dots < t_D$  denote the ordered event times and  $Z_{(i)k}$  be the  $k$ th covariate associated with the individual whose failure time is  $t_i$ . Define the risk set at time  $t_i$ ,  $R(t_i)$ , as the set of individuals who are still under study at a time just prior to  $t_i$ . Although the baseline hazard has not been specified in this partial likelihood we can apply maximum likelihood techniques.

Letting  $LL(\beta) = \ln[L(\beta)]$  and following some algebraic manipulation we can write  $LL(\beta)$  as

$$LL(\beta) = \sum_{i=1}^D \sum_{k=1}^p \beta_k Z_{(i)k} - \sum_{i=1}^D \ln \left[ \sum_{j \in R(t_i)} \exp\left(\sum_{k=1}^p \beta_k Z_{jk}\right) \right] \quad (17)$$

The (partial) maximum likelihood estimates are found by maximizing Equation (16), or, equivalently, (17).

A key reason for the popularity of the Cox model is that, even though the baseline hazard is not specified, reasonably good estimates of regression coefficients, hazard ratios of interest, and adjusted survival curves can be obtained for a wide variety of data situations. Another way of saying this is that the Cox PH model is a robust model, so that the results from using the Cox model will closely approximate the results for the correct parametric model.

## 6 Parametric models

Although the “robustness” provided by the non-parametric and semi-parametric techniques is attractive, these models cannot be used to extrapolate expected survival beyond the observed clinical trial data because of their lack of distributional assumptions.

Health economic evaluation is fundamentally interested in the mean costs and effects of treatment (Davies et al. 2012). As shown in Equation (2) the calculation of mean survival requires specification of the Survival function from  $t_0 \rightarrow t_\infty$ .

Because there are a large number of parametric models, I will focus my attention on the most common models. In the following sections, I will present their Survival and hazard functions. A description of the functional form of the hazard function will also be provided and what it entails for predicting/extrapolating expected survival.

### 6.1 Exponential Distribution

The simplest model for the hazard function is to assume that it is constant over time. The hazard of death at any time after the time origin of the study is then the same, irrespective of the time that has elapsed. Under this model, the hazard function may be written as

$$h(t) = \lambda$$

for  $0 \leq t < \infty$ .

From the definition of the survival function in (6) the corresponding survivor function is

$$\begin{aligned} S(t) &= \exp\left\{-\int_0^t \lambda dx\right\}, \\ &= \exp^{-\lambda t} \end{aligned}$$

Differentiating this result provides the pdf of the survival function as

$$f(t) = \lambda e^{-\lambda t} \tag{18}$$

for  $0 \leq t < \infty$ .

The constant hazard assumption means that exponential distribution is “memoryless”. For example, if the patient has survived nine months already, then memoryless means the probability that they will last another three months (so, a total of 12 months) is exactly the same as that of a newly affected patient lasting for the next three months.

## 6.2 Weibull Distribution

Because of the constant hazard assumption it is often quite difficult to justify the use of the exponential model for the long term extrapolation of survival probabilities. A more general form of hazard function is such that

$$h(t) = \lambda \gamma t^{\gamma-1} \quad (19)$$

for  $0 \leq t < \infty$ , a function that depends on two parameters  $\lambda$  and  $\gamma$ , typically known as the shape and scale respectively, which are both greater than zero. In the particular case where  $\gamma = 1$ , the hazard function takes a constant value  $\lambda$ , and the survival times have an exponential distribution. For other values of  $\gamma$ , the hazard function increases or decreases monotonically, that is, it does not change direction.

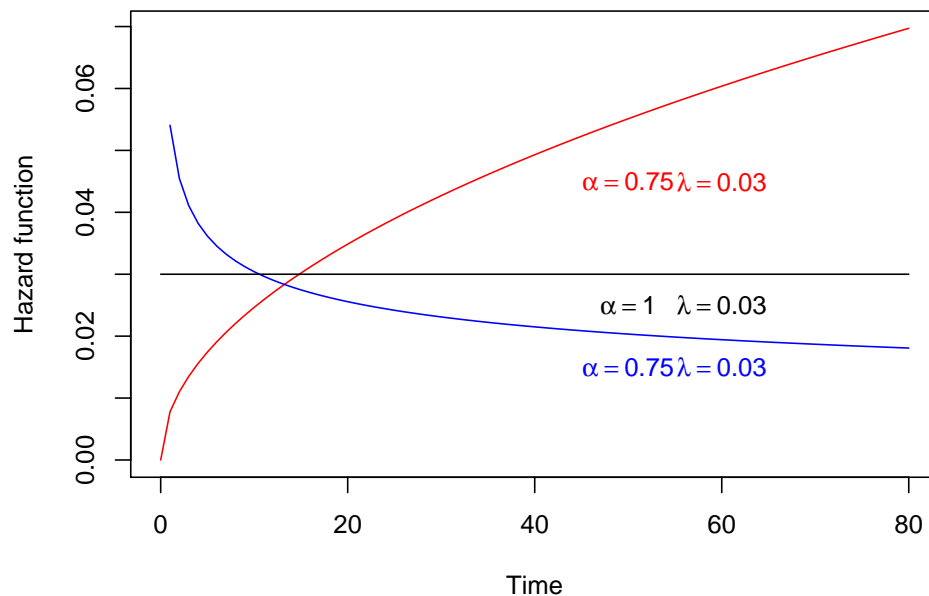


Figure 5: Weibull hazard functions

For this particular choice of hazard function, the survivor function is given by

$$\begin{aligned} S(t) &= \exp\left\{-\int_0^t \lambda \gamma x^{\gamma-1} dx\right\}, \\ &= \exp^{-\lambda t^\gamma} \end{aligned}$$

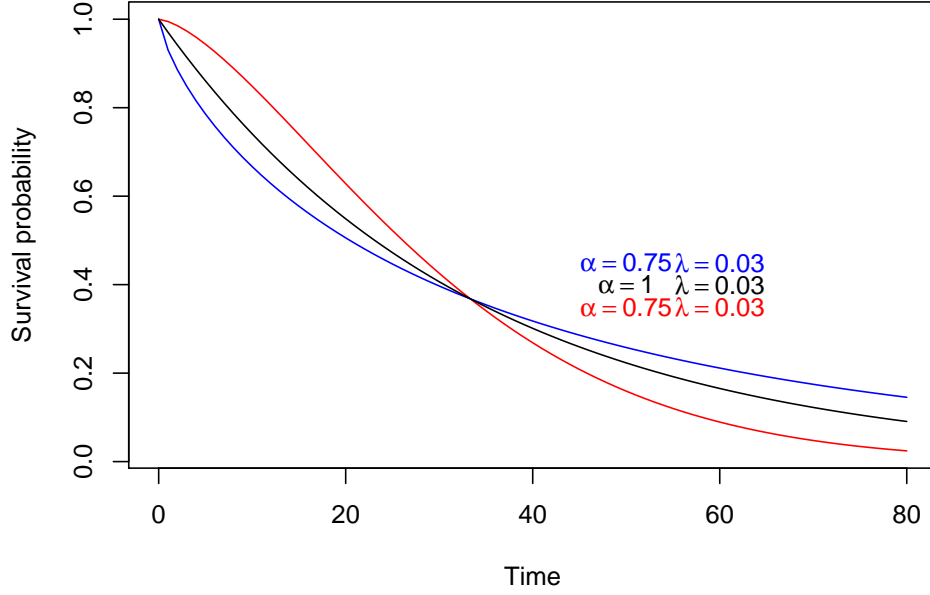


Figure 6: Weibull hazard functions

The corresponding probability density function is then

$$f(t) = \lambda \gamma t^{\gamma-1} \exp^{-\lambda t} \quad (20)$$

The additional flexibility introduced by the shape parameter makes the Weibull a more justifiable distribution for the extrapolation of the survival data particularly as it includes the exponential model as special case.

### 6.3 Gompertz Distribution

The Gompertz model is typically used in demography and has been found to be a good model for survival of adults in developed countries.

The hazard function of the Gompertz distribution is given by

$$h(t) = \lambda e^{\theta t} \quad (21)$$

for  $0 \leq t < \infty$ . In the particular case where  $\theta = 0$ , the hazard function has a constant value,  $\lambda$ , and the survival times then have an exponential distribution. The parameter  $\theta$  determines the shape of the hazard function, positive values leading to a hazard function that increases with time. The hazard function can also be expressed as  $h(t) = \exp(\alpha + \theta t)$ . The hazards of the Gompertz distribution increases or decreases monotonically.

The survivor function of the Gompertz distribution is given by

$$S(t) = \exp\left\{-\frac{\lambda}{\theta}(1 - e^{\theta t})\right\}$$

and the corresponding density function is

$$f(t) = \lambda e^{\theta t} \exp\left\{\frac{\lambda}{\theta}(1 - e^{\theta t})\right\}$$

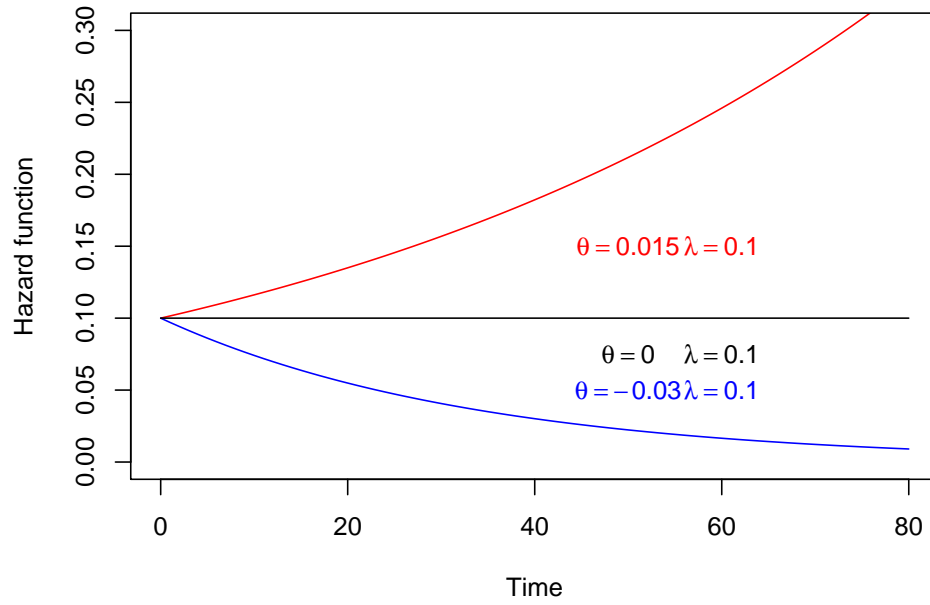


Figure 7: Gompertz hazard functions

In situations where the  $\theta > 0$ , the Gompertz model assumes exponentially increasing hazards which, dependent on the clinical disease area, may produce an overly pessimistic estimate of expected survival.

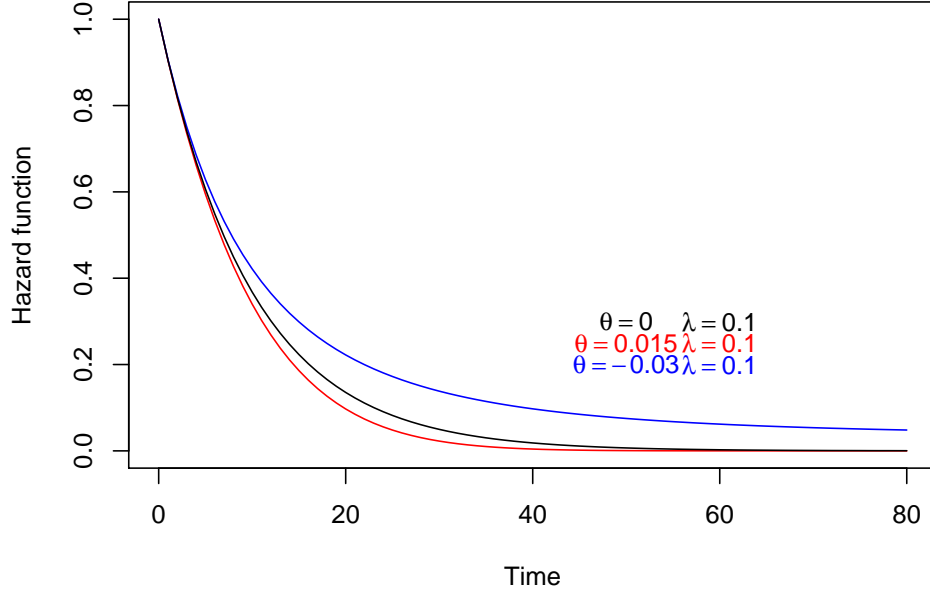


Figure 8: Gompertz Survival distributions

## 6.4 Log-Logistic Distribution

One potential limitation of the Weibull and Gompertz hazard functions is that they are a monotonic function of time. However, in some situations the hazard function may change direction over time. For example, following a heart transplant, a patient faces an increasing hazard of death over the first ten days or so after the transplant, while the body adapts to the new organ. The hazard then decreases with time as the patient recovers. In situations such as this, a unimodal hazard function may be appropriate.

A particular form of unimodal hazard is the function

$$h(t) = \frac{\alpha \lambda t^{\alpha-1}}{1 + \lambda t^\alpha} \quad (22)$$

$0 \leq t < \infty$ . The hazard function decreases monotonically if  $\alpha \leq 1$ , but  $\alpha > 1$ , the hazard has a single mode. The survival function is given by

$$S(t) = \{1 + \lambda t^\alpha\}^{-1} \quad (23)$$

and the probability density function is

$$f(t) = \frac{\alpha \lambda t^{\alpha-1}}{(1 + \lambda t^\alpha)^2} \quad (24)$$

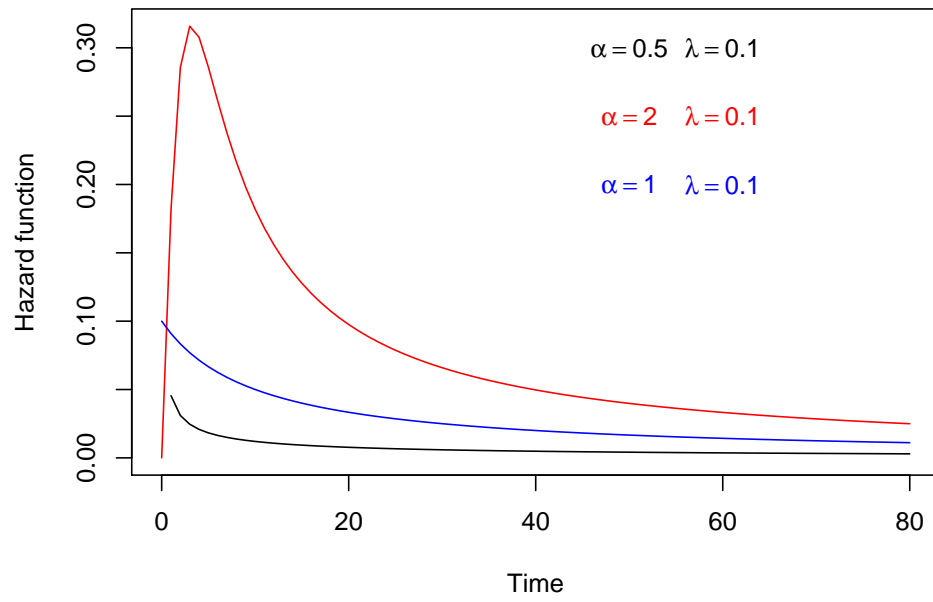


Figure 9: Log-Logistic hazard functions

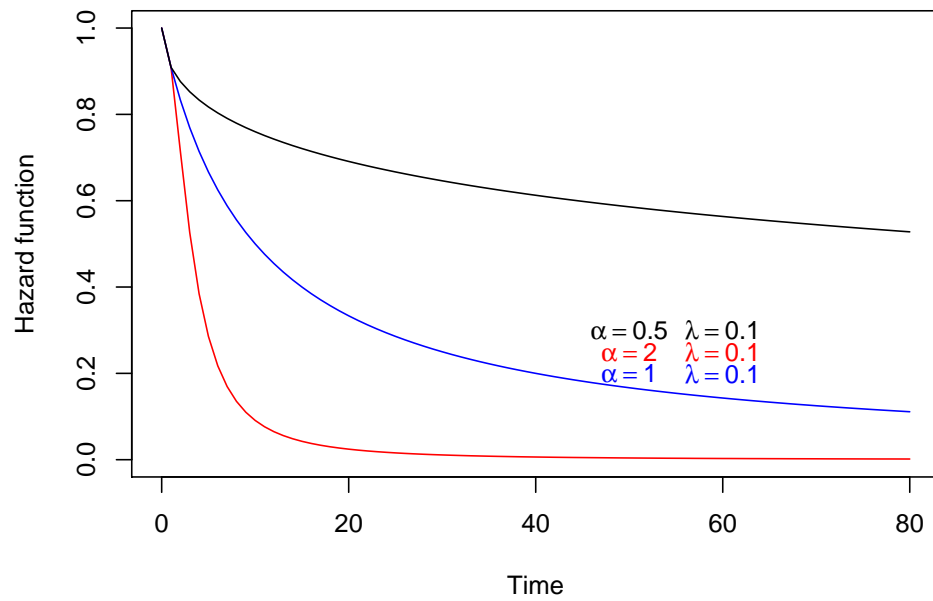


Figure 10: Log-Logistic Survival distributions



Although the non-monotonicity of the log-logistic distribution is a useful theoretical property, extrapolations based on this function typically result in substantial survival probabilities even as with large  $T$ . This is because after the mode of the hazard has been reached,  $h(t)$  tends to approach 0 as time approaches  $\infty$ .

#### 6.4.1 Note on parameterizations\*

It is important to note that different references, software packages and even packages within software programmes have different parameterizations of the survival functions (i.e. `survreg` vs `flexsurvreg`). One example is the parameterization of the Log-logistic distribution. The parameterization in `survreg` is the same as in `@(ref:illogis)`, with the coefficients referring to  $\mu$  and scale as  $\sigma$  (J. Klein and Moeschberger 2003).

In `flexsurvreg` a different parameterization is used which references Stata manual, although, the parameterization is not exactly the same, where  $b = \frac{1}{\lambda}$  and  $a = \frac{1}{\gamma}$ . In this formulation the hazard is decreasing for shape  $a \geq 1$ , and unimodal for  $a > 1$ . Therefore, before interpreting the results of an analysis it is prudent to verify the parameterization (and possibly manually plot the survival function to visually check if the results are sensible).

#### 6.4.2 Other Distributions

Many other survival distributions exist, and information on their hazard and survival functions can be found in (J. Klein and Moeschberger 2003). More complex three parameter models such as Generalized Gamma and Generalized F distributions offer even more flexibility, however the additional parameters make interpretation of their hazard functions more challenging.

Royston-Palmer models use cubic spline interpolation to create continuous hazard functions and between each changepoint (or “knot”) a separate Weibull or log-logistic model may be fit (Royston and Parmar 2002). These models have the theoretical advantage that they can fit any arbitrarily complex hazard with a sufficient number of changepoints. One disadvantage is that the location of knots is chosen arbitrarily, while a second more important potential disadvantage is the extrapolation is dependent on the parameters of the survival model in the last interval. If the number of events informing this interval is not sufficiently large there is the potential for the extrapolation to be confounded by random variation.

With the emergence of cell and gene therapies, there is a *suggestion* that some of these treatments could potentially offer a cure. Cure models assume that conditional on being treated a proportion of the patients are “non-susceptible” to the event while another group remain susceptible to the event  $p(x)$ .

$$S_{pop}(t|x, z) = 1 - p(x) + p(x)S_u(t|z) \quad (25)$$

In Cure models we may have introduced covariates  $X$  to influence the probability of being “susceptible” and (possibly different) covariates  $Z$  to influence the survival probabilities of the susceptibles. Some statistical methods have been proposed in order to test the hypothesis that the fraction observed at the end of a trial is indeed a cure fraction (Amico and Van Keilegom 2018), however, it may be unlikely that the clinical trial provides sufficient follow to use these methods.

## 7 Assessing the suitability of a parameteric model

Assessing whether a particular distribution for the survival times is plausible is to compare the survivor function for the data with that of a chosen model. A variety of diagnostic plots to assess the plausibility of different distributions are provided in the sections below.

## 7.1 Weibull or Exponential

For example, Suppose that a single sample of survival data is available, and that a Weibull distribution for the survival times is contemplated. Since the survivor function for a Weibull distribution, with scale parameter  $\lambda$  and shape parameter  $\gamma$ , is given by (??)

$$S(t) = \exp\{-\lambda t^\gamma\}$$

taking the logarithm of  $S(t)$ , multiplying by -1, and taking logarithms a second time, gives

$$\log(-\log S(t)) = \log \lambda + \gamma \log t \quad (26)$$

We now substitute the Kaplan-Meier estimate of the survivor function,  $\hat{S}(t)$  for  $S(t)$  in Equation (26). If the Weibull assumption is tenable,  $\hat{S}(t)$  will be “close” to  $S(t)$ , and a plot of  $\log\{-\log \hat{S}(t)\}$  against  $\log t$  would then give an approximately straight line.

If the log-cumulative hazard plot gives a straight line, the plot can be used to provide a rough estimate of the two parameters of the Weibull distribution.

Specifically, from Equation (26), the intercept and slope of the straight line will be  $\log \lambda$  and  $\gamma$ , respectively. Thus, the slope of the line in a log-cumulative hazard plot gives an estimate of the shape parameter, and the exponent of the intercept provides an estimate of the scale parameter.

Figure 11 demonstrates these results. The observations in the plot were simulated from a Weibull distribution with  $\lambda = 0.03$  and  $\gamma = 1$  (which is an exponential model). The  $R^2$  is almost 1 for the linear regression of plot, indicating the weibull is a very good fit for the data. Exponentiating the intercept and directly using the slope provide a reasonable approximation to the true parameters (0.034 & -3.373)

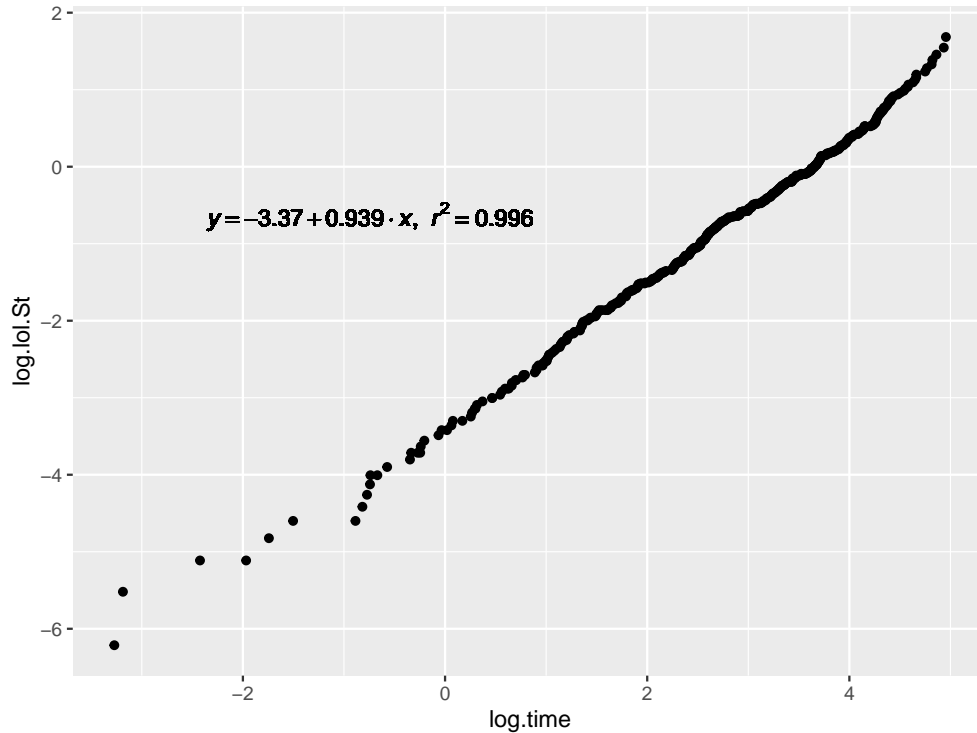


Figure 11: Diagnostic Plot for Weibull model

## 7.2 Gompertz

From Equation (21) it is clear to see that for a Gompertz distribution, the log hazards are linear with time. However, as discussed in Chapter 2, estimators for the hazard will tend to “jump around” or gyrate, especially when the time interval is small.

For example the Kaplan-Meier type estimate of the hazard would take the ratio of the number of deaths at a given death time to the number of individuals at risk at that time. If the hazard function is assumed to be constant between successive death times, the hazard per unit time can be found by further dividing by the time interval. Thus, if there are  $d_j$  deaths at the  $j$ th death time,  $t_{(j)}$ ,  $j = 1, 2, \dots, r$ , and  $n_j$  at risk at time  $t_{(j)}$ , the hazard function in the interval from  $t_{(j)}$  to  $t_{(j+1)}$  can be estimated by

$$\hat{h}(t) = \frac{d_j}{n_j \tau_j} \quad (27)$$

for  $t_{(j)} \leq t < t_{(j+1)}$ , where  $\tau_j = t_{(j+1)} - t_{(j)}$ .

Because the hazard would be estimated for an interval between each distinct event times, the hazard would gyrate very substantially. In order to provide a smoother hazard, the plot below presents the  $\log(\text{hazards})$  versus time, with the hazard assume to be equal for each interval (denoted by the horizontal lines). A regression for the  $\log(\text{hazard})$  for each interval vs median event time for that interval indicates an approximately linear relationship, which increases with larger samples sizes.

It is clear that this test is less conclusive than the test for weibull distribution 12, with highly linear plots being obtained from weibull models with large values for the shape ( $\gamma$ ) (Figure 13). However, for models the which are constant hazard, or which do not have rapidly increasing or decreasing hazards, no linear trend is observed in the plots (Figure 14).

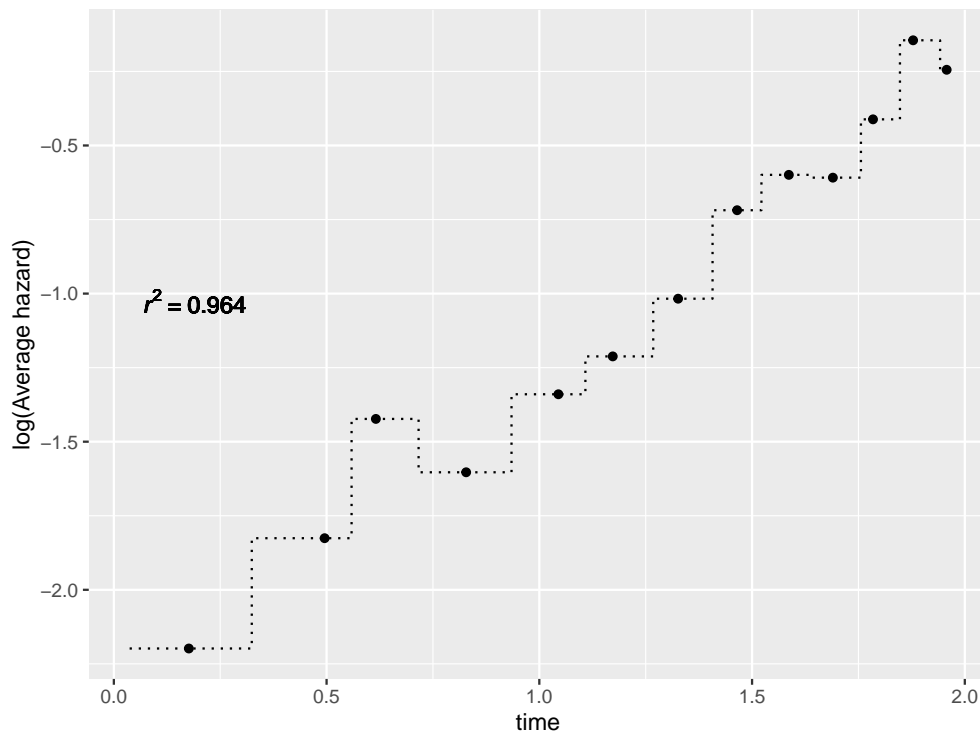


Figure 12: Diagnostic plot a gompertz when the underlying distribution is gompertz

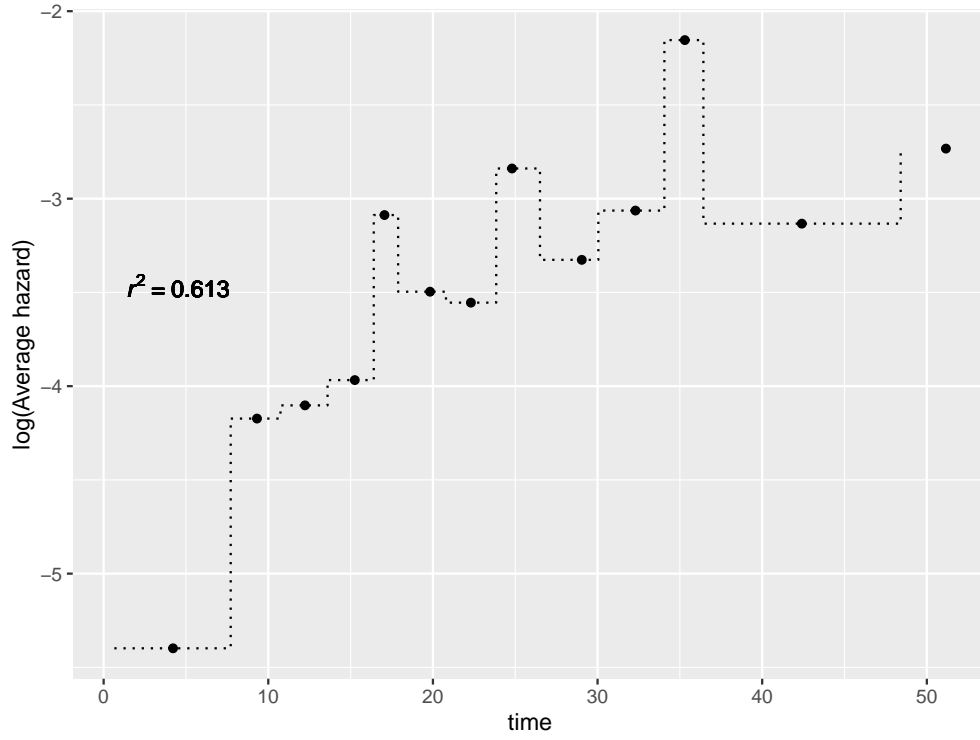


Figure 13: Diagnostic plot a gompertz when the underlying distribution is weibull with high value of shape

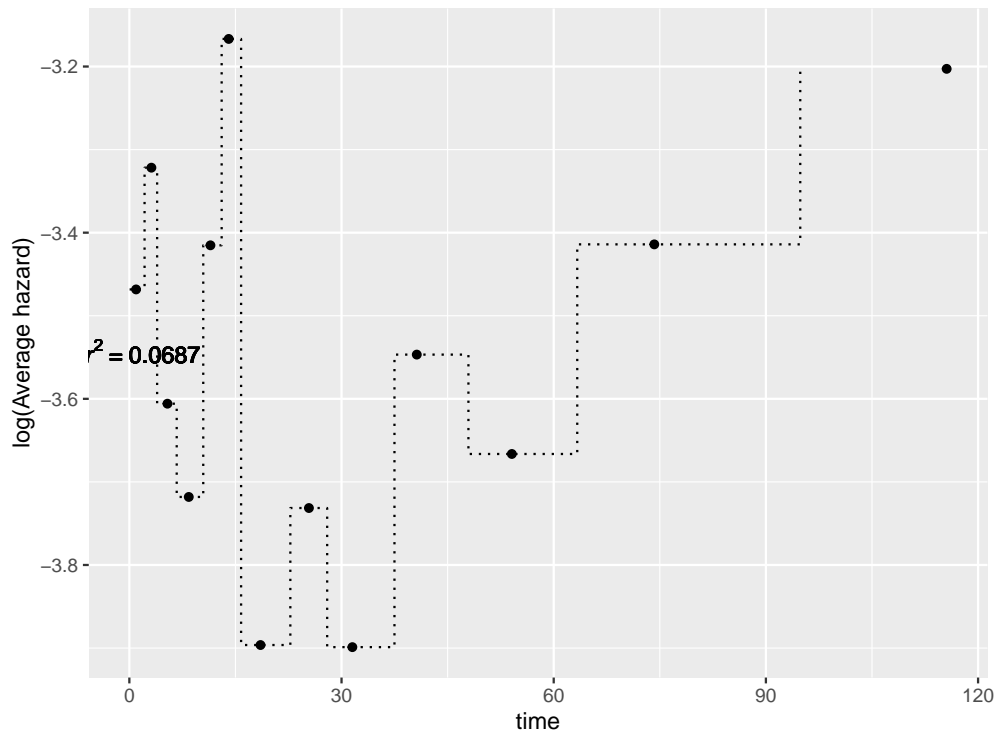


Figure 14: Diagnostic plot a gompertz when the underlying distribution is exponential

### 7.3 Log-logistic

Equation (23) presented the survival function of the log-logistic distribution. A distributional test for the log-logistic model can be derived from Survival/Failure Odds.

The survival odds is the odds of surviving beyond time  $t$  (i.e.,  $S(t)/(1 - S(t))$ ). This is the probability of not getting the event by time  $t$  divided by the probability of getting the event by time  $t$ .

The failure odds is the odds of getting the event by time  $t$  (i.e.,  $(1 - S(t))/S(t)$ ), which is the reciprocal of the survival odds. It is easy to show that the failure odds simplifies in a log-logistic model to  $\lambda t^\alpha$  (Klienbaum and Klein 2016).

Taking the natural logs, this can be rewritten as  $\log(\lambda) + \alpha[\log(t)]$ . In other words, the log odds of failure is a linear function of the log of time with slope  $\alpha$  and intercept  $\log(\lambda)$ .

The log-logistic assumption can be graphically evaluated by plotting  $\frac{\log(1-S(t))}{S(t)}$  against  $\log(t)$  where  $S(t)$  are the Kaplan–Meier survival estimates. If survival time follows a log-logistic distribution, then the resulting plots should be a straight line of slope  $\alpha$ .

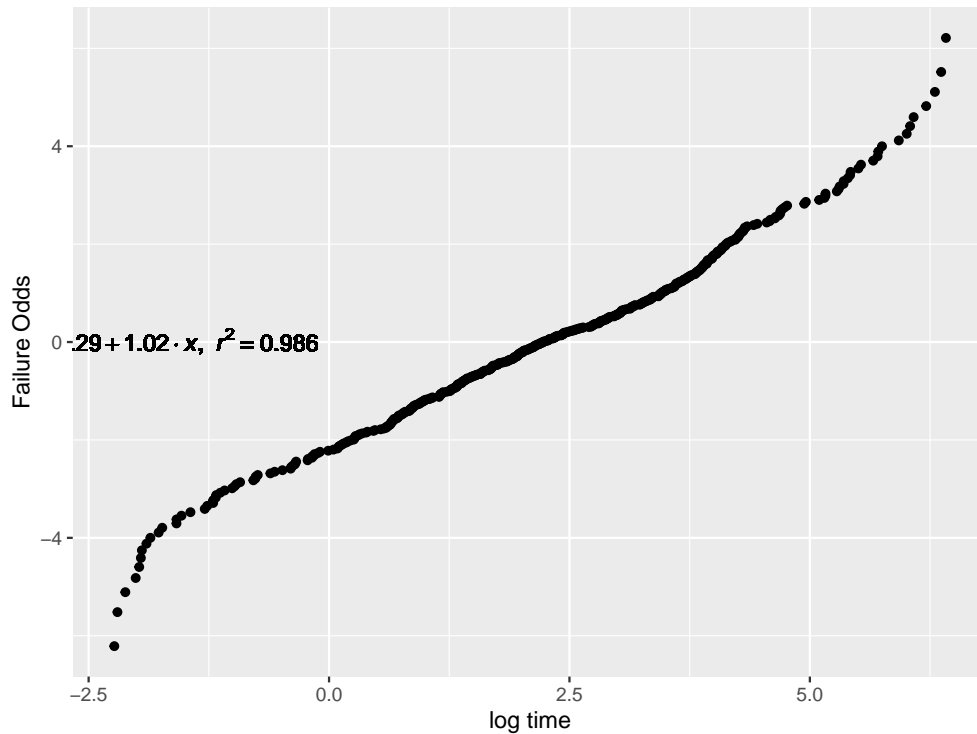


Figure 15: Diagnostic plot for log-logistic distribution

Figure 15 shows the diagnostic plot for a log-logistic distribution. The observations in the plot were simulated from a log-logistic distribution with  $\lambda = 0.1$  and  $\alpha = 1$ . The  $R^2$  is almost 1 for the linear regression of plot, indicating the log-logistic distribution is a very good fit for the data. Exponentiating the intercept and directly using the slope provide a reasonable approximation to the true parameters (0.101 & 1.019)

#### 7.3.1 Note

hat doesn't work with inline formatting in bookdown

## 8 Overview of Health Economics

The primary objective of this thesis is to investigate statistical techniques to inform for extrapolation of time to event data from clinical trials for the purposes of economic evaluation of therapies. Therefore, it is useful to provide a brief overview of health economics and more specifically - economic evaluation of medical interventions.

In many countries, health is primarily or substantially funded by the government. The basic idea of economic evaluation is to make a value judgement on a project (e.g. making a new pharmaceutical therapy publically available) involving public expenditure, given a finite budget and other potential investment choices. Other potential investment choices could be other medicinal therapies or more broadly additional health care staff or infrastructure.

Economic evaluation can be defined as a comparison of alternative options in terms of their costs and consequences (Drummond et al. 2005). Costs can be thought of as the value of the resources involved in providing a treatment or intervention; this would invariably include health care resources, and might be extended to include social care resources, those provided by other agencies, and possibly the time and other costs incurred by patients and their families or other informal carers. Consequences can be thought of as the health effects of the intervention. Two of the primary methods of economic evaluation are discussed in the subsequent sections.

### 8.1 Cost-effectiveness analysis (CEA)

In cost-effectiveness analysis we first calculate the costs and effects of an intervention and one or more alternatives, then calculate the differences in cost and differences in effect, and finally present these differences in the form of a ratio, i.e. the cost per unit of health outcome or effect (Weinstein and Stason 1977). Because the focus is on differences between two (or more) options or treatments, analysts typically refer to incremental costs, incremental effects, and the incremental cost-effectiveness ratio (ICER). Thus, if we have two options a and b, we calculate their respective costs and effects, then calculate the difference in costs and difference in effects, and then calculate the ICER as the difference in costs divided by the difference in effects:

$$ICER = \frac{Cost_a - Cost_b}{Effect_a - Effect_b} = \frac{\Delta Cost}{\Delta Effect}$$

The effects of each intervention can be calculated using many different types of measurement unit. Two diagnostic tests could be compared in terms of the cost per case detected, two blood pressure interventions by the cost per 1 mmHg reduction in systolic blood pressure, and two vaccination options by the cost per case prevented. However, decision-makers will typically be interested in resource allocation decisions across different areas of health care: for example, whether to spend more on a new vaccination programme or on a new blood pressure treatment. Consequently a measure of outcome that can be used across different areas is particularly useful, and the measure that has so far gained widest use is the quality-adjusted life-year (QALY).

### 8.2 Cost-benefit analysis (CBA)

As stated in the previous paragraph, a key distinction between CEA and CBA is the QALY. The QALY attempts to capture in one metric the two most important features of a health intervention: its effect on survival measured in terms of life-years, and its effect on quality of life.

The other key distinction is the concept of monetary value of health. CEA places no monetary value on the health outcomes it is comparing. It does not measure or attempt to measure the underlying worth or value to society of gaining additional health benefits, but simply indicates which options will permit more health benefits to be gained than others with the same resources.

In contrast, CBA attempts to place some monetary valuation on health outcomes as well as on health care resources. If a new surgical procedure reduces operative mortality by 5%, a cost-benefit approach would try to estimate whether each death averted had a value of €5000 or €500,000 or €5 million, and then assess whether the monetary value of the benefits was greater or less than the costs of obtaining these benefits.

Typically in Ireland figures of figures of €20,000 and €45,000 per QALY are defined as the willingness to pay threshold, i.e. how much the decision-maker is willing to spend to get an extra QALY (HIQA). Although budget constraints are an important consideration, Irish decision makers will typically fund new interventions that fall below this €45,000 per QALY threshold.

CBA allows for allocative efficiency, and (in theory) prioritize the reimbursement of different therapies across disease areas in terms of their cost-benefit ratio to make sure that the available resource are directed towards the therapies offering the largest health improvements.

### 8.3 Cost-effectiveness plane

When making assessments about CBA (in a slight abuse of notation simply referred to as cost-effectiveness), a new therapy will be compared against the current standard of care. This can be represented graphically in the form of the cost-effectiveness plane.

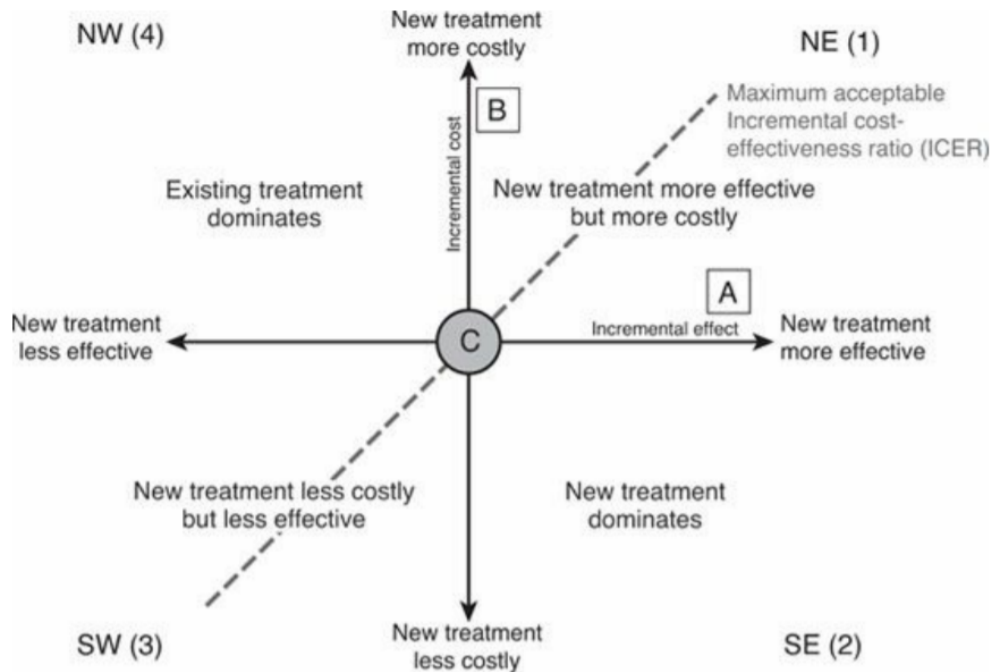


Figure 16: The cost-effectiveness plane

The most common situations arise in the north-east and south-west quadrants, where the new intervention is more effective but also more costly (the north-east quadrant, quadrant 1), or is less effective but also less costly (the south-west quadrant, or quadrant 3). In these areas of the figure, there is a trade-off between effect and cost: additional health benefit can be obtained but at higher cost (north-east), or savings can be made but only by surrendering some health benefit (south-west). In general if the ICER result is below the cost-effectiveness threshold it will be deemed cost-effective.

## 8.4 Valing Health - QALY

QALYs are a measure of outcome which typically assigns to each period of time a weight corresponding to the health-related quality of life during that period. Normally the weight 1 corresponds to full health and the weight 0 corresponds to a health state equivalent to dead. Figure 17 provides a graphical representation of the QALY approach, in which the life courses of two hypothetical individuals are plotted, with quality of life on the y-axis and time or survival on the x-axis. In this figure, both patients start with similar level of quality of life of 0.83 on a 0–1 scale. After approximately 1.5 years one patient has a complication ( $C^1$ ) which reduces her quality of life, and this is followed by three further complications ( $C^2$ ,  $C^3$ ,  $C^4$ ) which adversely affect quality of life, with the final one being fatal after approximately 7.5 years.

The second patient does not experience any complications until year 3 ( $C^5$ ), has one further non-fatal complication ( $C^6$ ), and then experiences a fatal complication just before year 9 ( $C^7$ ).

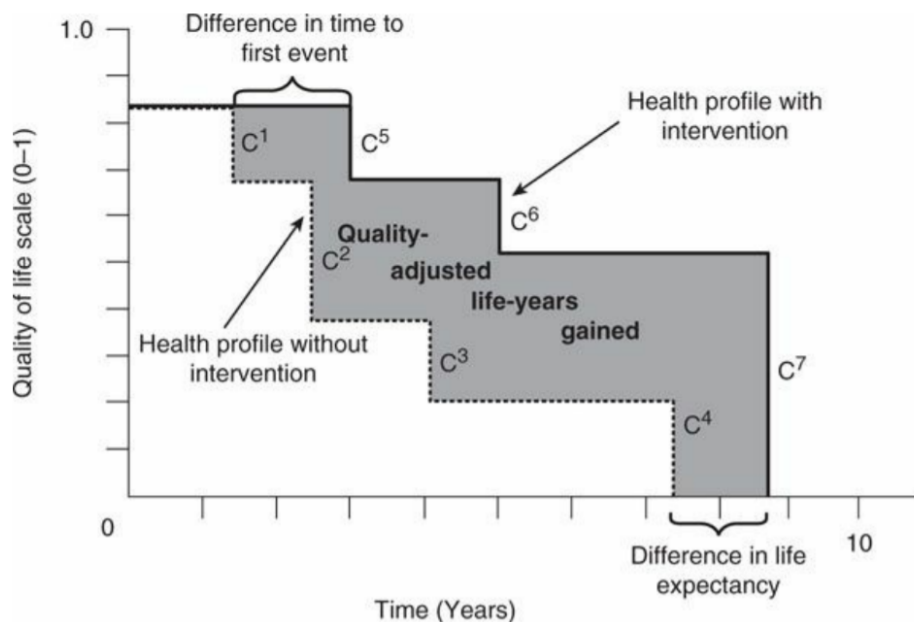


Figure 17: Health profile of two individuals in quality and quantity of life dimensions.

It can be seen that it would be possible to measure the difference between these two hypothetical patients in several different ways: by time to first event or complication-free time (a common measure in clinical trials), by time to death, or by number of complications. In this instance any of these would show some benefit to the patient receiving the intervention. However, all these are partial measures of the differences observed, and measuring the effect of the intervention using any single one of these metrics could be seriously misleading. In contrast, the area under each of the two curves or profiles does capture survival as well as the timing and number of non-fatal events and their health impact, and therefore the difference represented by the shaded area is a measure of QALYs gained. Parallels can be made to clinical trials in oncology, where, patients who are progression free have a certain level of quality of life, and which typically is reduced when they progress. Hence it is important to not only be able to estimate expected overall survival, but also expected progression free survival (or more generally any state of health that is expected to be meaningful different in terms of the QALY weight attached to it).

## 8.5 Cost effectiveness models

In order to use the information from a clinical to obtain cost-effectiveness results (i.e. ICER) a decision-analytic model (typically referred to as a cost-effectiveness model) is typically used. It has been noted that



relying on a randomized trial as a single vehicle for economic evaluation has a number of limitations (Sculpher et al. 2006). As a result, economic evaluation for decision-making will usually need to draw on evidence from a range of sources. These could include clinical, resource use, and outcome data collected alongside randomized trials, but are also likely to include evidence from other types of study such as cohort studies and surveys. A decision-analytic model provides a means of bringing together this full range of evidence and directing it at a specific decision problem being addressed by a health system at a given point in time and in a particular jurisdiction.

As per Drummond, cost-effectiveness (CE) models fulfill five main requirements of economic evaluation:

- Comparing all treatment options
- Reflecting all relevant evidence
- Linking intermediate to final end points
- Generalizing results to the decision-making context

and most importantly in the context of this PhD:

- Extrapolating over the appropriate time horizon of the evaluation

In summary, the cost-effectiveness model is the vehicle by which clinical trial data is combined with other information external to the trial, to obtain a cost-effectiveness result which is relevant to the jurisdiction of interest.

## 8.6 Expected Health Benefits and Costs

As stated in (M. Drummond 2015), expected value from a decision model represents the best estimate of the endpoints of interest for decision-making. It is the mean cost and effect, when multiplied by the number of patients treated, gives the total cost and overall health gain for that patient group (Davies et al. 2012) and therefore, the ICER is based on the Expected (or mean) Cost and QALYs.

This provides the primary motivation for this PhD research. As discussed in Section 6 and presented in Equation (2), calculation of the mean survival (to which we ascribe QALY weights) requires us to define the survival function for  $t_0 \rightarrow t_\infty$ . Because of censoring the full survival distribution is typically not available when making an assessment of the cost-effectiveness of a therapy. Therefore some type of extrapolation of the survival function is required in order to obtain the expected benefits of a treatment. Additionally, because time to treatment discontinuation may also be censored, extrapolation of this function is also required to obtain the expected costs of the intervention.

## References

Portions of the introduction to survival analysis have been taken from previously published materials and all credit is due to the respective authors. Figures 3 and 2 have been copied directly from (Collett 2015) along with their accompanying descriptions. Figure 4 has been copied from (Klienbaum and Klein 2016). Various other references such as (J. Klein and Moeschberger 2003), (Rizopoulos 2012) and (Moore 2016) were consulted during the creation of this review.

In the health economic section Figure 16 has been copied from (Gray 2012) along with much of the discussion around cost-effectiveness analysis. The section on cost-effectiveness modelling has been heavily borrowed

from (M. Drummond 2015).

Amico, Mailis, and Ingrid Van Keilegom. 2018. “Cure Models in Survival Analysis.” *Annual Review of Statistics and Its Application* 5 (1): 311–42. doi:10.1146/annurev-statistics-031017-100101.

Collett, D. 2015. *Modelling survival data in medical research*.

Cox, D. R. 1972. “Regression Models and Life-Tables.” *Journal of the Royal Statistical Society. Series B (Methodological)* 34 (2). [Royal Statistical Society, Wiley]: 187–220. <http://www.jstor.org/stable/2985181>.

Davies, Andrew, Andrew Briggs, John Schneider, Adrian Levy, Omar Ebeid, Samuel Wagner, Srividya Kotapati, and Scott Ramsey. 2012. “The Ends Justify the Mean: Outcome Measures for Estimating the Value of New Cancer Therapies.” *Health Outcomes Research in Medicine* 3 (February): e25–e36. doi:10.1016/j.ehrm.2012.01.001.

Drummond, Michael. 2015. *Methods for the economic evaluation of health care programmes*.

Gray, Alastair. 2012. *Applied methods of cost-effectiveness analysis in health care*. Oxford: Oxford University Press.

Klein, John, and Melvin L. Moeschberger. 2003. *Survival analysis : techniques for censored and truncated data*. Springer.

Klienbaum, Mitchel, and David G. Klein. 2016. *Survival Analysis : A self-learning text, Third edition*. Springer-Verlag New York.

Moore, Dirk F. 2016. *Applied survival analysis using R*. Springer.

Rizopoulos, Dimitris. 2012. *Joint models for longitudinal and time-to-event data : with applications in R*. Boca Raton, Fla.; London: Chapman & Hall/CRC.

Royston, Patrick, and Mahesh K. B. Parmar. 2002. “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* 21 (15): 2175–97. doi:10.1002/sim.1203.

Sculpher, Mark J, Karl Claxton, Mike Drummond, and Chris McCabe. 2006. “Whither trial-based economic evaluation for health care decision making?” *Health Economics* 15 (7): 677–87. doi:10.1002/hec.1093.