Estimation of Time Shift Models with Application to Survival Calibration in Health Technology Assessment

Andrew C. Titman

April 1, 2016

Abstract

The incremental life expectancy, defined as the difference in mean survival times between two treatment groups, is a crucial quantity of interest in cost-effectiveness analyses. Usually this quantity is very difficult to estimate from censored survival data with a limited follow-up period. The paper develops estimation procedures for a time shift survival model which, provided the model assumptions are met, gives a reliable estimate of incremental life expectancy without extrapolation beyond the study period. Methods for inference are developed both for individual patient data and when only published Kaplan-Meier curves are available. Through simulation the estimators are shown to be close to unbiased and constructed confidence intervals are shown to have close to nominal coverage for small to moderate sample sizes.

1 Introduction

Traditionally in survival analysis, mean survival is not considered an important summary measure (Hosmer & Lemeshow, p52) [1]. This is both because survival distributions are often heavily skewed making the mean survival a poor summary measure, and also because determining mean survival requires knowledge, or

This is the peer reviewed version of the following article:

Titman, A.C. (2016) Estimation of Time-Shift Models with Application to Survival Calibration in Health Technology Assessment. *Statistics in Medicine*. DOI:10.1002/sim.6951,

which has been published in final form at

`http://onlinelibrary.wiley.com/doi/10.1002/sim.6951/abstract`.

assumptions, about the whole survival distribution up to some maximum time horizon. However, obtaining estimates of mean survival is an integral part of the assessment of cost-effectiveness of health technologies. Specifically, cost-effectiveness analyses require estimation of incremental life expectancy, defined as the difference in mean survival times between treatment groups. Most trials and studies associated with an emerging health technology will have follow-up periods much shorter than the time horizon of interest meaning conclusions about cost-effectiveness may be sensitive to the method of survival extrapolation employed.

An approach which avoids these issues is to compare the restricted mean survival [2], which involves computing the mean of the minimum of an individual's survival time and a more limited time horizon, t^*. While the restricted mean survival can be reliably estimated from right-censored data provided the follow-up time exceeds t^*, there is no guarantee that differences in the restricted means will translate to similar differences in the incremental life expectancy. Cost-effectiveness is usually measured on the basis of incremental cost-effectiveness ratio (ICER) which depends upon the incremental life expectancy (or a quality-of-life adjusted version thereof). As a consequence, cost-effectiveness analyses usually require some degree of survival extrapolation.

Health Technology Assessments (HTA), such as those conducted by the UK's National Institute for Health and Care Excellence (NICE), involve building a decision-analytic model which encompasses the key events that may occur to patients and assesses their incidence under different treatments and the associated impact on cost, health utility and life expectancy. For tractability, a discrete-time Markov economic model is often adopted. This requires choices regarding the range of distinct health states and also the appropriate unit of time in the discrete time model (often referred to as the cycle length). In cancer trials or other assessments where there is a key irreversible intermediate event, such as progression or recurrence, which drives both quality of life and survival, a partitioned survival analysis approach is more commonly used. This involves estimating progression-free survival and overall survival separately and determining post-progression survival via the difference in these curves. In either type of approach it is usually necessary to make assumptions to allow extrapolation beyond the time span of the supporting studies.

An important concurrent validity check is whether the incremental life expectancy reported by the economic model is consistent with direct survival data from the key trial or trials. Substantial discrepancies between model based estimates and estimates from direct data will raise questions about the appropriateness of model assumptions or the sensitivity of the results to different forms of survival extrapolation.

Various approaches to survival extrapolation have been proposed in the literature. For instance, Latimer

[3] proposed a model selection algorithm based on assessment of the proportional hazards assumption and fitting a wide range of plausible parametric models to provide a sensitivity analysis to possible extrapolation assumptions. Demiris *et al* [4] considered use of a flexible class of poly-Weibull distributions in order to facilitate a wide range of hazard shapes including the 'bathtub' shaped hazards anticipated for organ transplantation studies. Bagust and Beale [5] argued against purely data driven approaches and against over-reliance on standard parametric models. One approach they advocate is the use of a time shift model, previously used in HTA reviews [6, 7]. If it is anticipated that the benefits of treatment will only last for a limited period of time it may be biologically plausible and also empirically justifiable to adopt such a model. The model has the advantage of providing an estimate of incremental life expectancy without requiring parametric assumptions about survival beyond the observed study period.

In this article, methodology based upon this time shift model is developed. In particular, formal estimation procedures are proposed and investigated, including construction of confidence intervals. The remainder of the article is as follows. Section 2 introduces the time shift model, Section 3 develops methods for estimation and construction of confidence intervals, Section 4 investigates the small sample properties of the estimators via simulation. In Section 5, the method is illustrated by an example relating to treatments for patients with BRAF V600E mutation melanoma. The paper concludes with a discussion.

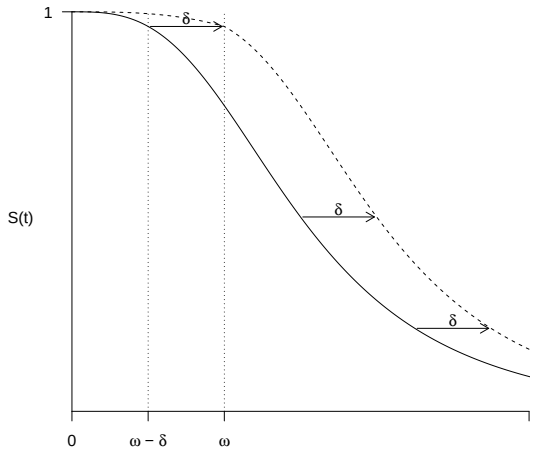
2 Time shift models

A useful potential model for survival data in contexts where incremental life-expectancy is of interest is a *time shift* model [5, 6, 7] where it is assumed that the survivor function for the group receiving the beneficial treatment converges to a shifted version of the survivor function for the group receiving standard treatment, after some initial period of survival benefit. We can characterize the model in terms of a time point ω, a time shift $\delta < \omega$, an initial beneficial treatment survival curve $\tilde{S}_2(t), 0 \leq t \leq \omega$ and the standard treatment survival curve $S_1(t)$. It is assumed that the overall survival for the beneficial survival group is

$$S_2(t) = \begin{cases} \tilde{S}_2(t) & \text{if } \ t < \omega \\ S_1(t-\delta) & \text{if } \ t \geq \omega. \end{cases}$$

For continuity it is necessary to assume $\tilde{S}_2(\omega) = S_1(\omega - t)$. A graphical representation of this model is given in Figure 1. The parameter ω represents the time by which all survival benefit of the new treatment will have accrued. Initially this quantity will be assumed to be known, but estimation will be considered in Section 3.4.

Figure 1: Graphical representation of a time shift model with a shift of δ and convergence point ω.



Note that general shift models of the form

$$F_2(t) = F_1(t + \delta)$$

have wide usage in statistics, not least in the arena of quantile regression where a linear model with respect to a particular quantile implies $q = F_2(t_q) = F_1(t_q + \delta)$ at the particular time point of interest, t_q where $t_q = F_2^{-1}(q)$. However, a uniform shift model where $S_2(t) = S_1(t + \delta)$ for all $t > 0$ is not usually appropriate for realistic survival models as it must imply a period of time of length greater than or equal to the shift δ in which no event is possible in the shifted group.

2.1 Estimation of incremental life-expectancy

The principal motivation for adopting a time shift model is to facilitate estimation of the incremental life expectancy in situations where there is a clinical or biological reason to suspect that the benefits of a new treatment will be accrued soon after commencement. The incremental life expectancy can be expressed in terms of the survival functions associated with the two treatments,

$$\Delta LE = \int_0^\tau \{S_2(t) - S_1(t)\}\, dt, \tag{2.1}$$

where τ is the maximum time horizon, which could theoretically be infinity and is usually much greater than the maximum follow-up time in available trial data.

Note that in most standard survival models, including proportional hazards and accelerated failure time models, in order to estimate (2.1) it is necessary to either make parametric assumptions or else have follow-up until the maximum time horizon. In contrast for the time shift model

$$\begin{aligned}\Delta LE &= \int_0^\tau \{S_2(t) - S_1(t)\}\, dt \\ &= \int_0^\omega \tilde{S}_2(t)dt + \int_\omega^\tau S_1(t-\delta)dt - \int_0^\tau S_1(t)dt \\ &= \int_0^\omega \tilde{S}_2(t)dt - \int_0^{\omega-\delta} S_1(t)dt,\end{aligned}$$

meaning that the incremental life-expectancy depends only on the follow-up to ω. Hence provided the follow-up time in available trial data exceeds ω, there is scope to estimate the quantity directly without any additional parametric assumptions. Moreover, in situations where $S_1(t) \approx 1$ for $t < \omega - \delta$ the shift, δ, itself can be used as a good approximation to the total incremental life expectancy.

3 Estimation of shift models

In this section methods for estimating and constructing confidence intervals for the shift parameter δ are developed. Initially the time by which all survival benefit has accrued, ω, is assumed to be known *a priori*. Note that, if the model is correctly specified, choosing any $\omega^* > \omega$ will still yield a valid estimate, but would be expected to give less efficient estimates. Subsequently in Section 3.4 methods for an unknown ω will also be considered.

3.1 Estimation of δ

A natural estimator of δ can be obtained simply by solving the equation

$$\hat{S}_1(\omega - \delta) = \hat{S}_2(\omega) \tag{3.1}$$

for δ, where $\hat{S}_j(t), j = 1, 2$ is the Kaplan-Meier estimator for group j. Hence

$$\hat{\delta} = \omega - \hat{S}_1^{-1}(\hat{S}_2(\omega)). \tag{3.2}$$

Clearly we would also like to determine whether δ is significantly different from 0 or some other null value and also obtain a 95% confidence interval for δ.

The estimator $\hat{\delta}$ can be defined in terms of an estimating equation

$$D(\delta) = \hat{H}_2(\omega) - \hat{H}_1(\omega - \delta) \tag{3.3}$$

where $D(\hat{\delta}) = 0$ and $E(D(\delta)) = 0$ and $\hat{H}_j(t) = -\log(\hat{S}_j(t))$ is the estimated cumulative hazard function. Through a simple application of Greenwood's formula and the assumption of independence between the two samples of patients an estimate of the variance of $D(\delta)$ can be obtained

$$\begin{aligned} \mathrm{Var}\{D(\delta)\} &= \mathrm{Var}\{\hat{H}_2(\omega)\} + \mathrm{Var}\{\hat{H}_1(\omega - \delta)\} \\ \widehat{\mathrm{Var}}\{D(\delta)\} &= \sum_{i:t_{2i} \le \omega} \frac{d_{2i}}{r_{2i}(r_{2i} - d_{2i})} + \sum_{i:t_{1i} \le \omega - \delta} \frac{d_{1i}}{r_{1i}(r_{1i} - d_{1i})} \end{aligned}$$

where t_{ji} denotes the ith ordered failure time in group j, and d_{ji} and r_{ji} are the number of failures and number at risk at the ith ordered event time for group j. The hypothesis test $H_0 : \delta = \delta_0$ against a general two-sided alternative can be conducted by noting that asymptotically, under the null hypothesis

$$\frac{D(\delta_0)^2}{\widehat{\mathrm{Var}}\{D(\delta_0)\}} \sim \chi^2_1.$$

Similarly, a $100(1-\alpha)\%$ confidence interval for δ can be constructed by considering the interval

$$\left\{ d \ : \ \frac{D(d)^2}{\widehat{\mathrm{Var}}\{D(d)\}} \le \chi^2_1(1-\alpha) \right\} \tag{3.4}$$

where $\chi^2_1(u)$ denotes the inverse CDF function of a χ^2_1 distribution. The performance of this construction for realistic sample sizes will be considered via simulation in Section 4.

Superficially, it may appear that an estimator based only on matching survival at ω would be inefficient. However, as shown in the Appendix, the estimator in (3.3) is in fact optimal among a class of estimators based on a weighted sum of such terms, at least in the case where $S_2(t)$ has a constant hazard for $t \ge \omega$.

A plug-in estimator of the incremental survival is given by

$$\hat{\Delta}LE = \int_0^{\omega} \hat{\tilde{S}}_2(t)dt - \int_0^{\omega - \hat{\delta}} \hat{S}_1(t)dt. \tag{3.5}$$

Unfortunately, while asymptotic theory for the standard error of restricted mean survival is relatively straightforward, see for instance Andersen *et al* (1993) page 279 [8] , the restriction that $\hat{\delta}$ satisfies $\hat{S}_1(\omega - \hat{\delta}) = \hat{\tilde{S}}_2(\omega)$ introduces a substantial complication. Approximate standard errors for ΔLE can be obtained through bootstrap resampling.

3.2 Estimation from published survival curves

In some circumstances, particularly if health economic modelling of the data involves secondary data analysis of previously published trials, it may not be possible to access the full, individual level, survival data. Often the data can be requested from the product sponsors, but if this is not possible then only the individual Kaplan-Meier curves plus the counts of numbers of patients at risk at a smaller range of time points, may be available.

Several authors have considered this issue in the context of fitting models for the purposes of meta-analysis [9, 10, 11]. The method of Guyot *et al* (2012) [9] allows estimates of the full set of sufficient statistics (numbers at risk and number of events) to be reconstructed meaning that the formulae given in Section 3.1 can then be applied. However, it is necessary to assume that the co-ordinates of the published survival curve can be extracted to sufficient accuracy to allow all event times can be uniquely identified.

Confidence intervals for δ based on (3.4) only require an estimate of $\widehat{\text{Var}}\{D(t)\}$. In Appendix 2, a method for estimating this quantity from published curves is developed, which is substantially simpler than attempting to reconstruct the full data and only requires an assumption that $\hat{S}_1(t)$ and $\hat{S}_2(t)$ can be determined for a fine grid of points, but not necessarily at all unique event times.

3.3 Goodness-of-fit and model sensitivity

A graphical check of the appropriateness of a time shift model can be obtained by calculating the optimal shift at a range of possible time points, ω. This corresponds to the *shift function*, $\hat{\delta}(t)$, considered for general shift models [12]. If the shift model is appropriate then plotting the estimated shift as a function of time should produce a plot that increases from time 0 until around the originally chosen ω and is then approximately flat thereafter. Pointwise 95% confidence intervals, computed using the interval in (3.4), can be included on the plot to aid the judgement.

If desired, and provided patient level data are available, further formal goodness-of-fit testing can be performed based on testing whether $S_1(t - \hat{\delta}) = S_2(t)$, for all $t > \omega$. The simplest way of doing this is to perform a log-rank test using the set of transformed survival times, $t_{1j} - (\omega + \delta)$ and $t_{2k} - \omega$. Note that since estimation of δ via (3.3) is independent of the survival beyond $\omega - \hat{\delta}$ and ω for groups 1 and 2, respectively, the log-rank test will have its standard asymptotic distribution. Alternative tests, for instance Renyi-type supremum tests [13], can also be used to detect deviations that are potentially non-proportional.

3.4 Estimation with unknown ω

It may not be possible or desirable to pre-specify a suitable value of ω. Estimation of δ in such cases is more complicated, but can still proceed through a two-step procedure based upon the estimated shift function.

Specifically the idea considered in Section 3.3 that, if a shift model is appropriate, the shift function should increase from $\hat{\delta}(0) = 0$ to reach $\hat{\delta}(\omega) = \delta_0$ and be constant thereafter.

Asymptotically the estimated shift at any point t is normally distributed about the true shift $\delta(t)$, where $\delta(t)$ satisfies

$$S_1(t - \delta(t)) = S_2(t)$$

for $t > 0$. This motivates the use of a weighted least-squares approach to jointly estimate δ_0 and ω. In order to proceed it is necessary to make stronger assumptions about the overall shift function than in previous sections. Previously the relationship between S_1 and S_2 for $t < \omega$ was left unspecified. Since it is guaranteed that $S_1(0) = S_2(0)$, the simplest possible specification is to assume the shift increases linearly until ω implying

$$\delta(t; \delta_0, \omega) = \left(\frac{t \wedge \omega}{\omega}\right) \times \delta_0,$$

where $t \wedge \omega = \min(t, \omega)$. This specification has the advantage of not requiring any additional nuisance parameters.

Estimation of δ_0 and ω can proceed by finding the (δ_0, ω) which minimize

$$\left\{\hat{\delta}(\mathbf{t}) - \delta(\mathbf{t}; \delta_0, \omega)\right\}' \mathbf{W} \left\{\hat{\delta}(\mathbf{t}) - \delta(\mathbf{t}; \delta_0, \omega)\right\}$$

where $\mathbf{t} = (t_1, \ldots, t_k)'$ is a set of times at which $\hat{\delta}(t)$ is estimated and $\mathbf{W}$ is a $(k \times k)$ positive-definite weight matrix. Choice of $\mathbf{W}$ affects the efficiency but not consistency of the estimator. The efficient weight matrix corresponds to the inverse of the variance-covariance matrix of $\hat{\delta}(\mathbf{t}) = (\hat{\delta}(t_1), \ldots, \hat{\delta}(t_k))'$. The theoretical covariance of $\hat{\delta}(t)$ at a pair of times (t, t')

$$\frac{\mathrm{Var}(\hat{H}_2(t \wedge t')) + \mathrm{Var}(\hat{H}_1(t^* \wedge t'^*))}{h_1(t - \delta(t)) h_1(t' - \delta(t'))},$$

where $t^* = t - \delta(t)$. However, this is difficult to reliably estimate due to the dependence on $h_j(t)$. Moreover the resulting matrix will be large and potentially ill-conditioned making inversion difficult.

A pragmatic alternative which avoids these problems while maintaining reasonable efficiency, is to use a diagonal weight matrix with terms

$$W_{kk} = \left[\widehat{\mathrm{Var}}\{\hat{H}_2(t_k)\} + \widehat{\mathrm{Var}}\{\hat{H}_1(t_k - \hat{\delta}(t_k))\}\right]^{-1},$$

where $\widehat{\text{Var}}(\hat{H}_j(t))$ is defined as in (6.3).

A plug-in estimate of the incremental life expectancy from the model can be found via

$$\hat{\Delta}LE = \int_0^{\hat{\omega}} \hat{S}_1(t - \delta(t; \hat{\delta}_0, \hat{\omega}))dt - \int_0^{\hat{\omega}-\hat{\delta}_0} \hat{S}_1(t)dt, \tag{3.6}$$

where $\hat{S}_1(t)$ is the Kaplan-Meier estimator of survival for the control group. Standard error estimates for $\hat{\omega}, \hat{\delta}_0$ and $\hat{\Delta}LE$ can be obtained via bootstrap resampling.

Note that there are some non-standard properties to the model. In particular, under the null where $\delta_0 = 0$, the convergence point ω is not identifiable. Similarly if the shift function is linear or close to linearly increasing or decreasing throughout the follow-up period then $\hat{\omega}$ may be unbounded. In this case, a convention that $\hat{\omega} = t_k$ is proposed.

4 Simulation study

In this section, the finite sample properties of the proposed estimators are investigated, both for individual patient data and from published Kaplan-Meier estimates, using simulation.

A shift model is simulated by assuming patients in the control group have exponential survival times, such that $S_1(t) = \exp(-\lambda t)$, while the experimental group has

$$S_2(t) = \exp[-\lambda\{t - (t \wedge \omega)/\omega\}].$$

The rate parameter λ is chosen so that mean survival in the control group is 5 years, The true value of ω is taken to be 2 months and $\delta = 0.5$ years (183 days). Two censoring scenarios are considered each producing 50% follow-up within one year; exponential censoring with rate $\log(2)/10$ and uniform censoring from a $U[0, 20]$ distribution. Sample sizes of $N = 100, 200, 500$ and 1000 per group are considered.

In order to investigate the effect of misspecifying ω on the estimates of δ and the incremental life-expectancy, ΔLE, estimates of $\hat{\delta}$ using the estimating equation in (3.3) are obtained under four scenarios where ω is assumed to be 2, 1.5, 3 and 4 months, respectively. In each case, $\hat{\Delta}LE$ is also obtained by applying the estimator in (3.5). In addition, to investigate the properties of the estimators proposed in Section 3.4, δ and ΔLE are also estimated assuming ω is unknown.

The results for estimates of δ are summarized in Table 1. In the scenario considered, there is evidence of some negative bias in the estimated shift δ for small to moderate sample sizes when ω is correctly specified. As we would expect, there is little impact on bias, but a reduction in efficiency if a larger value of ω is

Table 1: Bias and Standard Deviation (SD) of estimates of the shift δ (measured in days) for different sample size and censoring scenarios and different assumptions about ω.

		Bias					SD				
N	Censoring	$\omega = 2$	$\omega = 1.5$	$\omega = 3$	$\omega = 4$	$\hat{\omega}$	$\omega = 2$	$\omega = 1.5$	$\omega = 3$	$\omega = 4$	$\hat{\omega}$
100	Uniform	-6.39	-50.02	-4.43	-6.81	67.47	157.99	134.02	218.18	278.59	294.91
100	Exponential	3.07	-45.55	1.52	-10.96	68.41	155.55	132.56	222.39	279.24	310.35
200	Uniform	-2.15	-45.28	-2.75	-4.14	54.10	111.07	93.40	154.63	197.08	215.28
200	Exponential	-2.94	-44.69	-3.44	-2.02	50.87	112.88	92.60	157.29	200.75	214.66
500	Uniform	-0.78	-46.34	-2.29	1.62	32.83	69.65	58.18	97.25	124.80	125.69
500	Exponential	-0.94	-45.23	0.17	-1.04	33.95	71.00	60.02	97.25	123.74	126.07
1000	Uniform	0.44	-45.72	-0.17	-0.12	16.32	50.27	42.13	67.86	87.04	81.53
1000	Exponential	0.56	-46.30	-0.42	0.63	16.92	49.89	41.69	68.60	88.57	85.96
2000	Uniform	-0.27	-45.51	-0.45	-0.32	2.66	22.00	18.86	30.27	39.41	31.20
2000	Exponential	0.13	-45.64	-0.15	-0.75	2.15	22.01	18.75	30.45	39.39	31.83

Table 2: Bias and Standard Deviation (SD) of estimates of the incremental life expectancy ($\hat{\Delta}LE$) (measured in days) for different sample size and censoring scenarios and different assumptions about ω.

		Bias					SD				
N	Censoring	$\omega = 2$	$\omega = 1.5$	$\omega = 3$	$\omega = 4$	$\hat{\omega}$	$\omega = 2$	$\omega = 1.5$	$\omega = 3$	$\omega = 4$	$\hat{\omega}$
100	Uniform	-5.03	-38.63	-2.32	-4.09	39.23	138.37	121.39	169.89	194.34	198.07
100	Exponential	3.40	-34.52	2.52	-5.40	38.42	136.77	120.19	173.21	194.15	203.42
200	Uniform	-1.44	-34.47	-2.08	-2.81	28.66	97.83	84.94	121.12	138.09	142.34
200	Exponential	-2.15	-33.94	-2.40	-0.78	27.11	99.12	84.03	123.23	140.35	142.16
500	Uniform	-0.33	-35.55	-1.29	1.39	16.37	61.12	52.81	76.44	88.31	84.45
500	Exponential	-0.60	-34.53	0.23	0.24	17.11	62.34	54.46	76.26	86.43	84.56
1000	Uniform	0.69	-35.05	0.06	0.32	7.85	44.23	38.25	53.37	61.30	57.06
1000	Exponential	0.64	-35.56	-0.22	0.41	7.78	43.81	37.90	53.71	62.19	59.14
2000	Uniform	-0.08	-34.83	-0.23	-0.05	1.62	19.36	17.15	23.78	27.71	24.83
2000	Exponential	0.25	-34.94	0.00	-0.39	1.10	19.36	17.03	23.86	27.61	25.32

Table 3: Empirical coverage of nominal 95% confidence intervals under different assumptions regarding reporting of number of patients at risk assuming $\omega = 2$ months is known.

N	Censoring	Cov_1	Cov_2	Cov_3
100	Uniform	0.953	0.951	0.952
100	Exponential	0.944	0.943	0.944
200	Uniform	0.947	0.946	0.947
200	Exponential	0.952	0.952	0.952
500	Uniform	0.950	0.950	0.950
500	Exponential	0.951	0.949	0.950
1000	Uniform	0.953	0.952	0.952
1000	Exponential	0.945	0.944	0.945
2000	Uniform	0.948	0.948	0.948
2000	Exponential	0.952	0.952	0.952

chosen. Choosing too small a value of ω leads to persistent underestimation of the shift parameter. When ω is treated as unknown there is a substantial positive bias in the estimates of δ for small to moderate sample sizes. This is due to instances where the estimated shift function has a linear trend that persists for the whole period of follow-up leading to positively biased estimates of ω.

For estimates of the incremental life expectancy (ΔLE), given in Table 2, there is a broadly similar pattern to the direction and magnitude of bias in the cases where ω is assumed known. However, it is noticeable that there is less bias in ΔLE than δ when ω is treated as unknown.

Additionally, in the case where $\omega = 2$ months is assumed known, 95% confidence intervals are constructed using individual patient data (Cov_1), the Kaplan-Meier curves plus numbers at risk every 0.5 years (Cov_2) or every 1 year (Cov_3). Table 3 gives the empirical coverage of the constructed 95% confidence intervals in each case. Confidence intervals derived from intermittently reported numbers at risk displayed very little difference from those based on individual patient data, with the discrepancy decreasing for large sample sizes.

5 Example: Survival of patients melanoma with BRAF V600E mutation

As an illustrative example, the time shift model is applied to data on patient survival in a randomized controlled trial of patients with unresectable, previously untreated stage IIIc or stage IV melanoma that tested positive for the BRAF V600E mutation. Patients were randomized to either vemurafenib or dacarbazine. The original study demonstrated a statistically significant survival benefit in treatment by vemurafenib [14]. Subsequently, approval for use in the National Health Service in England and Wales was sought via a NICE Single Technology Appraisal (STA). Survival follow-up in the original study was only 10 months [14], but updated data extending this to 18 months was available for the STA. Nevertheless, in order to estimate the mean life-years gained between treatments considerable survival extrapolation was necessary.

Survival extrapolation based on an assumption of a shared constant hazard beyond 9 months, produces an estimated survival difference of around 3.41 months. However, the authors incorporated external evidence from the dacarbazine arm of a previous trial of melanoma patients with unknown BRAF mutation status [15] which suggested a decreasing hazard for longer term survivors treated with dacarbazine. Once these considerations were incorporated the estimated survival difference increased around threefold, with a corresponding improvement in cost-effectiveness.

The Evidence Review Group's report noted that shifting the Kaplan-Meier curve of dacarbazine by approximately 97 days, meant it lined up with the curve for vemurafenib and hence deduced that the gain from vemurafenib was also approximately 97 days. This approach can be formalized by applying the methods of Section 3.2 based on the data from Figure 13 of the company's report [16], which gives the number of patients at risk at monthly intervals. Taking the time of convergence, ω, to be 5 months gives an estimated shift of 3.13 months (95% CI: 2.31, 3.47). Figure 2 shows the Kaplan-Meier curves with the curve for dacarbazine shifted. There is seen to be good agreement between the shifted dacarbazine survival and the vemurafenib survival after 5 months.

The good fit of the shift model is further confirmed through scrutiny of the shift function which remains stable after 5 months (Figure 3).

If the length of time, ω, before benefit was accrued is not assumed known it is necessary to use the least-squares method proposed in Section 3.4. Using this method, taking time points at 0.1 month intervals up to 15 months gives an estimated shift of 3.22 months (95% CI: 2.17, 4.89) occurring from 4.24 months (95% CI: 3.04, 7.17).

Figure 2: Kaplan-Meier estimates of survival for vemurafenib and dacarbazine groups, with the estimated shift of $\delta = 3.13$ months at convergence time $\omega = 5$ months.

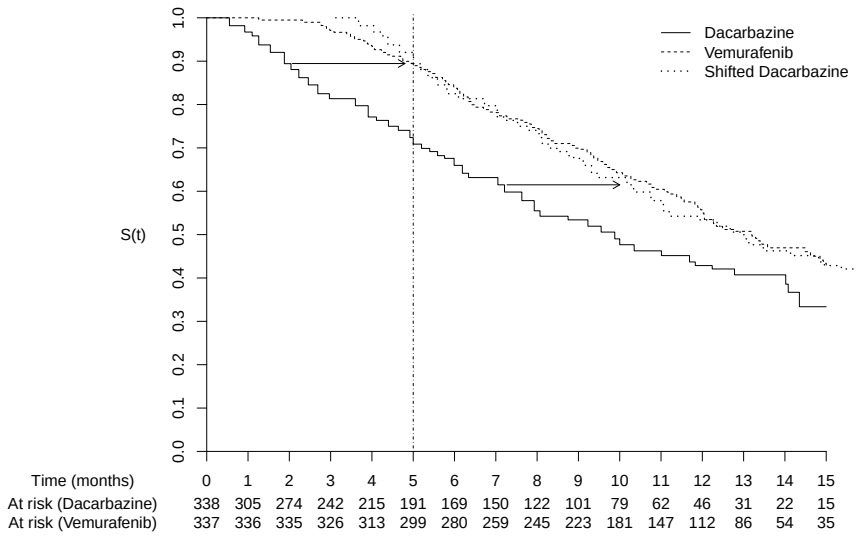
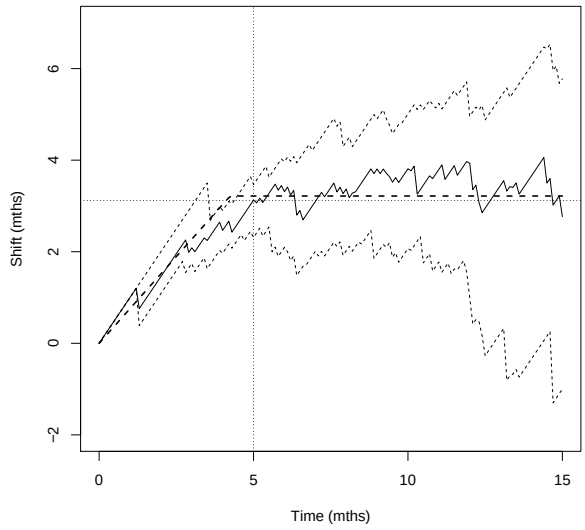


Figure 3: Estimated shift function with pointwise 95% confidence limits for the BRAF V600E mutation data. Bold dashed line represents the estimated shift assuming an unknown ω.



The incremental survival, $\hat{\Delta}LE$, obtained using (3.5) is estimated as 3.04 months (95% CI: 2.44, 3.63) if $\omega = 5$ is assumed known and 3.17 months (95% CI: 2.21, 4.54) if ω is estimated from the data.

The time shift model fits the available direct trial data very closely. In addition, unlike other models which could be applied, it produces an incremental life expectancy estimate that is independent of particular assumptions about the shape of the long-term hazard in the dacarbazine arm. In particular, it is still consistent with the possibility of a long-term decreasing hazard for both groups. In the absence of information about long-term survival, the assumption used in the time shift model that the hazards in the two treatment arms are the same up to the specified fixed shift cannot be empirically tested. Therefore a definitive estimate of the difference in mean survival between treatments cannot be obtained. However, the estimate from the shift-model does raise considerable doubt about the credence of more favourable estimates based on other methods which also required, potentially stronger, untestable assumptions about long-term survival.

6 Discussion

The time shift model is useful in situations where it is expected that the benefits of a treatment will be short-term. It is straightforward to efficiently estimate the shift and obtain valid 95% confidence intervals. Good statistical properties can be demonstrated if ω can be pre-specified in advance. There is also scope to jointly estimate ω and δ from the data, but much larger sample sizes are required to ensure estimates are close to unbiased.

A limitation of the time shift model is that it does not give a direct estimate of the mean survival in each individual treatment group. Moreover, it does not fully indicate at what time points the difference in mean survival is accrued. This restricts the direct use for cost-effectiveness analysis, where some form of discounting is usually applied, meaning survival accrued far in the future is given lower weighting. For these quantities it is still necessary to make some assumptions about the form of $S_1(t)$ beyond the range of the data. In practice, the majority of the benefit occurs by time ω and within the follow-up period, making the results relatively insensitive to any sensible choice for the long-term hazard. The restriction that $S_1(t - \delta) = S_2(t)$ can also be built into a more general decision-analytic model.

The shift-model should be considered as a possible approach to survival extrapolation in an HTA, but only if it is biologically plausible and there is empirical evidence of its appropriateness. Assuming biological plausibility, a practical strategy would be to first scrutinize the shift-function; if this shows evidence of the characteristic flattening required then we would proceed to formally estimate the shift model using the

methods in Section 3. If there is no evidence of convergence then alternative extrapolation methods, such as fitting exponential or other parametric models to the tails of the respective survival curves on the basis of the trend in the cumulative hazard functions [5], should be used. Such an approach should yield estimates of incremental life expectancy that are less prone to bias than the procedure used in Section 4 where the shift-model was applied regardless of evidence of a convergence point within the follow-up period. However, fully formalizing this approach and therefore assessing its statistical properties requires additional work.

In the BRAF V600E example considered, the shift model represents a very good fit to the data. Subsequent clinical experience also seems to have confirmed its appropriateness, with acknowledgement that most patients develop a resistance to vemurafenib within 6-7 months [17]. Similarly, in the application of the model in an STA relating to non-small-cell lung cancer the apparent time-shift observed in the overall survival was supported by additional information on post-progression survival requested by the Evidence Review Group which revealed the new treatment provided no post-progression survival benefit [7].

In other situations, even if the underlying process does broadly adhere to such a pattern and even if the sample size of the study is reasonably large, there may be variation away from this pattern particularly towards the end of the follow-up period. The necessity to have empirical justification for the shift model is likely to preclude its use in studies with small sample sizes. The simulation results suggests that the approach should be avoided or the estimates viewed with caution for data with less than 200 patients per treatment group. In such cases it may be safer to fix a conservatively large ω, rather than attempt to estimate it, in order to limit the possibility of bias.

Acknowledgements

The author is grateful to Professor Adrian Bagust for useful discussions on the practical utility of the time shift model and also to the associate editor and two anonymous referees for their helpful and constructive comments.

References

[1] Hosmer, D.W., Lemeshow, S. (1999). *Applied Survival Analysis: Regression Modelling of Time to Event Data* Wiley, New York.

[2] Royston, P., Parmar, M. K. (2011). The use of restricted mean survival time to estimate the treatment effect in randomized clinical trials when the proportional hazards assumption is in doubt. *Statistics in Medicine*. 2011, **30**: 2409-2421.

[3] Latimer, N.R. Survival analysis for economic evaluations alongside clinical trialsextrapolation with patient-level data inconsistencies, limitations, and a practical guide. *Medical Decision Making*. 2013, **33**: 743-754.

[4] Demiris N, Lunn D., Sharples L.D. Survival extrapolation using the poly-Weibull model. *Statistical Methods in Medical Research*, 2015, **24:** 287-301.

[5] Bagust A, Beale S. Survival Analysis and Extrapolation Modeling of Time-to-Event Clinical Trial Data for Economic Evaluation: An Alternative Approach. *Medical Decision Making*, 2014, **34:** 343-351.

[6] Beale S, Dickson R, Bagust A, Blundell M, Dundar Y, Boland A, Marshall E, Plummer R, Proudlove C. Vemurafenib for the treatment of locally advanced or metastatic BRAF V600 mutation-positive malignant melanoma: a NICE single technology appraisal. *Pharmacoeconomics*, 2013, **31:** 1121-1129.

[7] NICE *Pemetrexed maintenance treatment following induction therapy with pemetrexed and cisplatin for non-squamous non-small-cell lung cancer* Technology appraisal guidance documents. 2014. Available at `http://www.nice.org.uk/guidance/TA309/documents/lung-cancer-non-small-cell-non-squamous-pemetrexed-maintenance-following-pemetrexed-cisplatin-evaluation-report4`

[8] Andersen, P.K., Borgan Ø, Gill R.D., Keiding N. *Statistical Models Based on Counting Processes*. 1993, Springer, New York.

[9] Guyot, P., Ades, A. E., Ouwens, M. J., Welton, N. J. Enhanced secondary analysis of survival data: reconstructing the data from published Kaplan-Meier survival curves. *BMC Medical Research Methodology*, 2012, **12** (1), 9.

[10] Wan, X., Peng, L., Li, Y. A Review and Comparison of Methods for Recreating Individual Patient Data from Published Kaplan-Meier Survival Curves for Economic Evaluations: A Simulation Study. *PLoS one*. 2015; **10**(3).

[11] Hoyle MW, Henley W. Improved curve fits to summary survival data: application to economic evaluation of health technologies. *BMC Medical Research Methodology*. 2011; **11**:139.

[12] Doksum, K. Empirical Probability Plots and Statistical Inference for Nonlinear Models in the Two-Sample Case. *Annals of Statistics*, 1974, **2:** 267-277.

[13] Fleming, T. R., Harrington, D. P., O'Sullivan, M. Supremum versions of the log-rank and generalized Wilcoxon statistics. *Journal of the American Statistical Association*. 1987, **82:** 312-320.

[14] Chapman PB, Hauschild A, Robert C, Haanen JB, Ascierto P, Larkin J, et al. Improved survival with vemurafenib in melanoma with BRAF V600E mutation. *New England Journal of Medicine*. 2011, **364:** 2507-2516.

[15] Robert C, Thomas L, Bondarenko I, O'Day S, Weber J, Garbe, C *et al* Ipilimumab plus dacarbazine for previously untreated metastatic melanoma. *New England Journal of Medicine*. 2011, **364**: 2517-2526.

[16] Roche Products Limited. *NICE STA submission: Vemurafenib for the treatment of locally advanced or metastatic BRAF V600 mutation positive malignant melanoma (TA269)*. 2012. Available at `http://www.nice.org.uk/guidance/ta269/documents/melanoma-braf-v600-mutation-positive-unresectable-metastatic-vemurafenib-roche-products4`

[17] Swaika A. Crozier JA, Joseph RW. Vemurafenib: an evidence-based review of its clinical utility in the treatment of metastatic melanoma. *Drug Design, Development and Therapy*. 2014, **8**: 775-787.

Appendix 1: Efficiency of the estimator for δ

In this appendix it is shown that the shift estimator based on the estimating equation in (3.3) is the most efficient amongst a class of weighted estimators in the case where the hazard functions are constant after time ω.

The estimating equation in (3.3) can be extended to a case where δ is chosen based on the weighted discrepancy in shifted cumulative hazards at a series of evaluation times $t_j \geq \omega$ such that

$$D^*(\delta) = \sum_{j=1}^{J} w(t_j) \left\{ \hat{H}_2(t_j) - \hat{H}_1(t_j - \delta) \right\}, \tag{6.1}$$

for some weight function $w(t) \geq 0$. The asymptotic variance of an estimator $\hat{\delta}$ satisfying $D^*(\delta) = 0$ is given by $\mathrm{Var}(D^*(\delta_0))/E(\frac{\partial \bar{D}(\delta_0)}{\partial \delta})$ where $\bar{D}(\delta)$ is (6.1) with $\hat{H}_j(t)$ replaced by $H_j(t)$ for $j = 1, 2$ and δ_0 is the true

shift. It is convenient to note that $D^*(\delta)$ can be rewritten in terms of increments of the estimated cumulative hazard, which are independent. Specifically

$$D^*(\delta) = \sum_{j=1}^{J} w^*(t_j) \left(\Delta\hat{H}_2(t_j) - \Delta\hat{H}_1(t_j - \delta) \right)$$

where $w^*(t_j) = \sum_{k=j}^{J} w(t_j)$, $\Delta\hat{H}_k(t_1) = \hat{H}_k(t_1)$ and $\Delta\hat{H}_k(t_j) = \hat{H}_k(t_j) - \hat{H}_k(t_{j-1})$ for $j > 1$. Under this representation it can be shown that

$$\mathrm{Var}(D^*(\delta)) = \sum_{j=1}^{J} w^*(t_j)^2 \left\{ \mathrm{Var}(\Delta\hat{H}_2(t_j)) + \mathrm{Var}(\Delta\hat{H}_1(t_j - \delta)) \right\}$$

and

$$E\left\{ \frac{\partial \bar{D}(\delta_0)}{\partial \delta} \right\} = w^*(t_1)h_2(t_1) + \sum_{j=2}^{J} w^*(t_j) \left\{ h_2(t_j) - h_2(t_{j-1}) \right\}. \tag{6.2}$$

In the case where the hazard function $h_2(t)$ is constant beyond ω, all the terms in the summation in (6.2) are equal to zero. It therefore follows that minimizing $\mathrm{Var}(\hat{\delta})$ is achieved by setting $w^*(t_j) = 0$ for $j > 1$. In addition, since $\mathrm{Var}(\hat{H}_j(t))$ are increasing functions in t, it follows that the most efficient consistent estimate is achieved when $t_j = \omega$. In other situations it may be possible to get slightly more efficient estimators through judicious use of weighting, but this would require knowledge of the shape of the hazard function and also the censoring distributions.

Appendix 2: Estimation of $\mathrm{Var}(D(\delta))$ from published survival curves

In this section, a simplified method for obtaining an estimate of $\mathrm{Var}(D(\delta))$ from published survival curves is developed.

It is assumed that $\hat{S}_j(t)$ is known at a set of points $\{t_{jl}, l = 1, \ldots, L\}$ but, unlike the approach of Guyot *et al* [9], these do not necessarily coincide with the exact set of event times. In addition it is assumed that the number of patients at risk in group j, $R_j(t)$, is known at a set of points $\{u_{jk}, k = 1, \ldots, K\}$, which is a subset of $\{t_{jl}, l = 1, \ldots, L\}$.

For mathematical convenience it is assumed that the number of events at any event time is small compared to the number at risk such that the asymptotic variance estimator for the Nelson-Aalen estimate of the cumulative hazard,

$$\widehat{\mathrm{Var}}(\hat{H}_j(t)) = \sum_{i:t_{ji} \leq t} \frac{d_{ji}}{r_{ji}^2}, \tag{6.3}$$

is appropriate.

It is then assumed that the hazard function $h_j(t)$ is piecewise constant between consecutive evaluation times t_{l-1j}, t_{lj} and that the censoring hazard $g_j(t)$ is constant between consecutive evaluation times of the number at risk, u_{k-1j}, u_{kj}.

A natural estimator for h_{lj}, the constant hazard assumed between times t_{l-1j} and t_{lj} is then

$$\frac{\log\{\hat{S}_j(t_{l-1j})\} - \log\{\hat{S}_j(t_{lj})\}}{t_{lj} - t_{l-1j}}.$$

Similarly since $R_j(t) = R_j(0)S_j(t)G_j(t)$ where $G_j(t)$ is the survivor function of the censoring distribution, a natural estimator for g_{kj}, the constant censoring hazard assumed between times, u_{k-1j} and u_{kj} is then

$$\frac{\log\{\hat{S}_j(u_{k-1j})R_j(u_k)\} - \log\{\hat{S}_j(u_{kj})R_j(u_{k-1j})\}}{u_{kj} - u_{k-1j}}.$$

Based on this construction, an estimate of the number at risk at the evaluation times t_{lj} can be established via

$$\hat{R}_j(t_{lj}) = R_j(0)\hat{S}_j(t_{lj})\hat{G}_j(t_{lj})$$

where

$$\hat{G}_j(t_{lj}) = \exp\left\{ - \sum_{k:u_{kj}<t_{l+1j}} g_{kj}(\min(u_{kj}, t_{lj}) - u_{k-1j}) \right\},$$

which is the survivor function resulting from the estimated piecewise constant censoring hazards.

The increment in (6.3) between evaluation times, t_{l-1j} and t_{lj} can then be estimated by

$$\begin{aligned} \hat{I}_{lj} &= \frac{1}{\hat{R}_j(t_{l-1j})} \int_{t_{l-1j}}^{t_{lj}} h_{lj} \exp\{(g^*_{lj} + h_{lj})(t - t_{l-1j})\} dt \\ &= \frac{h_{lj}\left[\exp\{(g^*_{lj} + h_{lj})(t_{lj} - t_{l-1j})\} - 1\right]}{\hat{R}_j(t_{l-1})(g^*_{lj} + h_{lj})} \end{aligned}$$

where $g^*_{lj} = g_{k^*j}$ where k^* is the unique k satisfying $u_{k-1j} \le t_{lj} < u_{kj}$.

The estimate for $\mathrm{Var}\{D(d)\}$ to be used in (3.4) is then

$$\widehat{\mathrm{Var}}\{D(d)\} = \sum_{i:t_{2i}\le\omega} \hat{I}_{2i} - \sum_{i:t_{2i}\le\omega-d} \hat{I}_{1i}.$$