ELSEVIER

Controlled Clinical Trials 24 (2003) 682–701

Controlled Clinical Trials

On the use and utility of the Weibull model in the analysis of survival data

Kevin J. Carroll, M.Sc.*

AstraZeneca Pharmaceuticals, Biostatistics Group, Global Clinical Science, Alderley Park, Macclesfield, UK

Manuscript received September 9, 2002; manuscript accepted March 17, 2003

Abstract

In the analysis of survival data arising in clinical trials, Cox's proportional hazards regression model (or equivalently in the case of two treatment groups, the log-rank test) is firmly established as the accepted, statistical norm. The wide popularity of this model stems largely from extensive experience in its application and the fact that it is distribution free—no assumption has to be made about the underlying distribution of survival times to make inferences about relative death rates. However, if the distribution of survival times can be well approximated, parametric failure-time analyses can be useful, allowing a wider set of inferences to be made. The Weibull distribution is unique in that it is the only one that is simultaneously both proportional and accelerated so that both relative event rates and relative extension in survival time can be estimated, the latter being of clear clinical relevance. The aim of this paper is to examine the use and utility of the Weibull model in the analysis of survival data from clinical trials and, in doing so, illustrate the practical benefits of a Weibull-based analysis.

Keywords: Survival data; Proportional hazards regression; Weibull model; Hazard ratio; Event time ratio

Introduction

Cox's proportional hazards regression model (or equivalently in the case of two treatment groups, the log-rank test) has become the statistician's mainstay in the analysis of survival data [1–6]. Its predominance stems from 3 decades of application and experience, together with the fact that it is distribution free; no assumption has to be made about the underlying distribution of survival times to make inferences about relative death rates. While this is a key

* Corresponding author: Kevin Carroll, AstraZeneca Pharmaceuticals, Biostatistics Group, Global Clinical Science, Alderley Park, Macclesfield, SK10 4TG, UK. Tel.: +44-1625-515234; fax: +44-1625-518537.

E-mail address: kevin.carroll2@astrazeneca.com

0197-2456/03/$—see front matter
doi:10.1016/S0197-2456(03)00072-2

strength of the model, it does introduce some limitations. Specifically, a direct quantification of the improvement in survival time is not possible, except in the special case of truly exponentially distributed lifetimes where the reciprocal of the hazard ratio estimates the ratio of median times to event [7]. However, lifetimes are seldom truly exponential in their distribution, so statisticians have tended to rely on Kaplan-Meier estimates of the underlying survivor function to read off estimated percentiles. The reliability and precision of these estimates depends upon the number of deaths and patients remaining at risk at any given time point on the curve. Median survival time is often used as a measure of improvement in time, though this measure is often unavailable at the earlier analyses of longer-term trials with relatively low event rates. Even when median survival can be estimated from the Kaplan-Meier curve, tests for differences in medians between treatments are generally approximate and do not directly link with tests for parity of hazard rates [8].

The Weibull model provides an alternative, fully parametric approach to the Cox model. Both these models are, in fact, closely related; both assume proportional hazards and both provide asymptotically unbiased, equally efficient estimates of the hazard ratio between two treatments. The Weibull model, in addition to being proportional, is simultaneously an accelerated failure-time model (AFT), and is the only parametric distribution to possess both properties [4,9]. AFT models simply examine survival times via a log-linear model so that treatment effects are expressed in terms of the relative increase or decrease in survival time. The Weibull, being both accelerated and proportional, therefore allows the simultaneous description of treatment effects both in terms of hazard ratios and also in terms of the relative increase or decrease in survival time; we might conveniently refer to this latter quantification of treatment effect as an "event time ratio," if only to illustrate the close parallel with the better known hazard (or event) rate ratio. Cox has suggested that these kinds of analyses are most favorable when a direct interpretation of the treatment effect is desired [10].

It is important to recognize that the Weibull and other AFT models are not new, having previously been described in the literature [11]. A good, accessible overview can be found in Colette [4]. Prentice and Kalbfleisch [9] and Wei [12] have discussed the potential use of AFT models in survival analyses and, more recently, Chen and Wang [13,14] have discussed AFT models alongside a new class of models, the "accelerated hazards model," which models how the underlying hazard changes over time.

Despite such coverage in the literature, the Weibull model is rarely used in the routine analysis and reporting of clinical trial data. Given that the Weibull allows simultaneous estimation of both the usual hazard ratio and an event time ratio, in addition to allowing a more thorough examination of proportionality and providing a means for predicting how data might mature over time, further consideration of its use and usefulness seems worthwhile.

The remainder of this paper is therefore structured as follows: the next section provides an overview of the Weibull model, including its form, estimation of hazard and event time ratios, examination of proportionality, and prediction of data maturation. After this a comparison of Cox and Weibull models in the analysis of real clinical trial data is made, followed by a brief discussion on the need for an exact distributional match when using the Weibull model. A brief summary of key results is then followed by the final section discussing the practical value and application of the Weibull and related models in the analysis of survival data in arising clinical trials.

The Weibull model

Before describing the Weibull model, it is helpful to consider a general distribution for lifetimes for which proportionality holds.

Let $T = t$ denote the time to some event of interest; this could be time to death or progression-free survival in an oncology setting. If $f(t)$ denotes the probability density function of T, $S(t)$ the survivor function, and $h(t)$ the hazard function, then, as is well known,

$$f(t)=h(t)e^{-\int_0^t h(u)du}$$

Under proportionality, $h_A(t) = \theta h_B(t)$, so that $S_A(t) = [S_B(t)]^\theta$, where θ is the hazard ratio and A and B denote two independent treatment groups.

The maximum likelihood estimate of the hazard ratio is the easily derived:

$$\hat{\theta}_{para} = e^{\hat{\gamma}_{para}} = \frac{\sum_{N_B}\int_0^{t_i} h(u)du}{\sum_{N_A}\int_0^{t_i} h(u)du}\,\frac{d_A}{d_B} \tag{1}$$

and

$$\hat{V}(\hat{\gamma}_{para}) = \frac{1}{d_A} + \frac{1}{d_B} \tag{2}$$

where d_A and d_B denote the total number of deaths observed in treatment groups A and B, respectively. Full details of this result are given in the appendix.

Armitage and Berry give an estimate of the hazard ratio associated with the Cox (log-rank) model [6],

$$\hat{\theta}_{\mathrm{Cox}} = e^{\hat{\gamma}_{\mathrm{Cox}}} = \frac{d_A}{E_A}\frac{E_B}{d_B} = \frac{d_A}{\sum r_{iA}\frac{d_i}{r_i}}\,\frac{\sum r_{iB}\frac{d_i}{r_i}}{d_B} \tag{3}$$

where, at time t_i, there are a total of d_i events out of r_i subjects at risk with d_{iA} events out of r_{iA} at risk in group A and d_{iB} events out of r_{iB} at risk in group B so that $E(d_{iA}) = (r_{iA})\, d_i/r_i$; and d_A and d_B denote the total number of deaths in groups A and B, respectively. If γ is small then,

$$\hat{V}(\hat{\gamma}_{\mathrm{Cox}}) = \frac{1}{d_A} + \frac{1}{d_B} \tag{4}$$

Under the assumption of proportionality, Eqs. (2) and (4) show that the standard error for the log hazard ratio is asymptotically the same for *all* underlying distributions, $f(t)$, and is the same as that for the Cox model estimate. Thus, the use of parametric analyses does not lead to any asymptotic loss of efficiency compared to the log-rank or Cox analysis under the assumption of proportionality [2,9]. Furthermore, upon close examination of Eqs. (1) and (3), we can see that, under the assumption of proportionality, both quantities are estimating the average risk of death on treatment A relative to that on B. Hence, any parametric analysis where proportionality is assumed to hold, such as the Weibull (or simpler exponential), will give rise to an estimated hazard ratio very similar to that from a conventional Cox analysis.

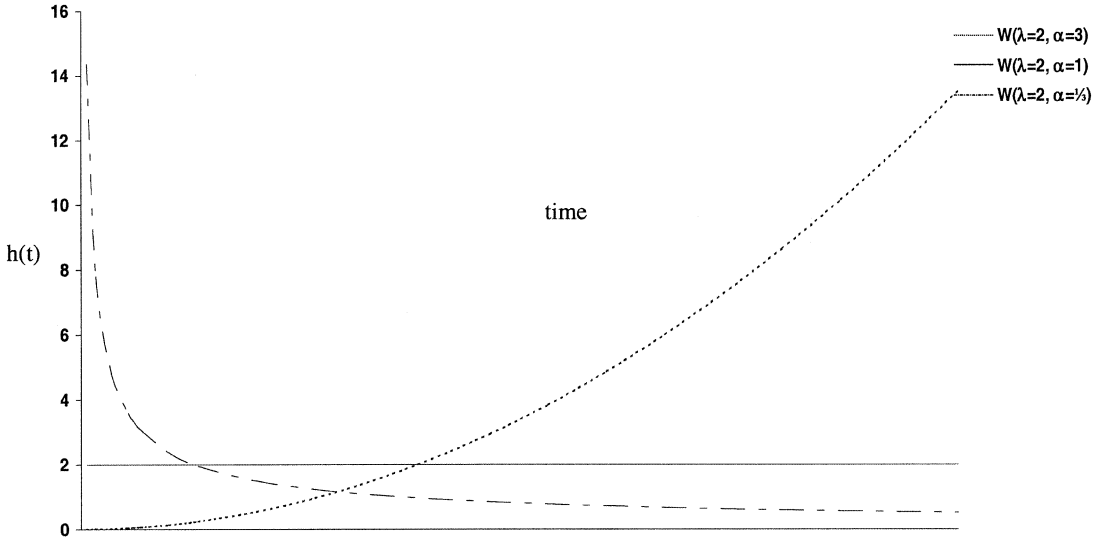

Fig. 1. The Weibull hazard function.

In general, therefore, we should not be concerned in employing parametric models, such as the Weibull, when proportionality holds.

Now, returning to the specifics of the Weibull model. If T represents the time-to-event variable, then the probability density function of the Weibull distribution is given by

$$f(t) = \alpha\lambda t^{\alpha-1} e^{-\lambda t^{\alpha}} \tag{5}$$

where $\lambda > 0$ is the event rate parameter and $\alpha > 0$ is the scale, or shape parameter. Thus, $S(t) = e^{-\lambda t^{\alpha}}$ and $h(t) = \alpha\lambda t^{\alpha-1}$. Note that the variable $Y = T^{\alpha}$ is a simple exponential with parameter λ. An illustration of $h(t)$ is given in Fig. 1. An alternative parameterization of the Weibull is given by setting

$$\alpha = \frac{1}{\sigma}, \quad \text{and} \quad \lambda_i = e^{-(\mu+\underline{\beta}'\underline{x}_i)/\sigma} \tag{6}$$

where the influence of the covariates, $\underline{x}_i$, for the ith individual is modeled through the event rate parameter, λ_i. (The software package SAS uses this parameterization when fitting the Weibull in the procedure PROC LIFEREG [15].) We shall now consider the important features of this distribution.

The hazard ratio

Based on the parameterization in Eq. (5), the hazard ratio for two treatments is given by

$$\theta(t) = \frac{\alpha_A\lambda_A}{\alpha_B\lambda_B} t^{\alpha_A-\alpha_B} \tag{7}$$

Hence, if $\alpha_A \neq \alpha_B$, hazards are not proportional. When proportionality does hold, $\theta = \lambda_A/\lambda_B$.

Based on the parameterization in Eq. (6), the log hazard ratio for an individual with covariates $\underline{x}_i$ relative to an individual with covariates $\underline{x}_j$ is

$$\frac{-\underline{\beta}'(\underline{x}_i - \underline{x}_j)}{\sigma} \tag{8}$$

In the case of just two treatments, the log hazard ratio is therefore given by $-\beta/\sigma$. In SAS the variance of the estimated log hazard ratio, $-\hat{\beta}/\hat{\sigma}$, is not given directly but can be easily derived from the variance-covariance matrix via Taylor's expansion [6],

$$\hat{V}\left(-\frac{\hat{\beta}}{\hat{\sigma}}\right) \cong \left(\frac{\hat{\beta}}{\hat{\sigma}}\right)^2 \left\{\frac{V(\hat{\beta})}{\hat{\beta}^2} - \frac{2Cov(\hat{\beta},\hat{\sigma})}{\hat{\beta}\hat{\sigma}} + \frac{V(\hat{\sigma})}{\hat{\sigma}^2}\right\}$$

As was noted above, under proportionality a Weibull analysis will give rise to an estimated hazard ratio and standard error very similar to that obtained from a conventional Cox analysis. This close matching of outcomes is easily verified by simulation. Table 1 shows the results of a simple simulation study where deviates from a range of Weibull distributions were randomly generated and analyzed in SAS by Cox's proportional hazards regression and also by assuming a Weibull distribution. For each Weibull shown, 1000 datasets were simulated, for sample sizes of 250, 100, and 25 in each of two treatment groups. A random amount of censoring (10%) was incorporated. A hazard ratio of 0.8 was used throughout.

Table 1. Simulated Weibull data: analysis by Cox and by Weibull

			Cox analysis			Weibull analysis		
n[a]	Shape	Event rate on treatment A	HR[b]	5th and 95th percentiles	SE[c] log HR	HR	5th and 95th percentiles	SE log HR
250	$\alpha = 1/3$	$\lambda_A = 0.5$	0.801	0.679, 0.934	0.0991	0.802	0.681, 0.935	0.0983
		$\lambda_A = 2$	0.801	0.685, 0.938	0.0966	0.800	0.686, 0.935	0.0955
	$\alpha = 3$	$\lambda_A = 0.5$	0.803	0.692, 0.937	0.0949	0.804	0.692, 0.937	0.0946
		$\lambda_A = 2$	0.796	0.685, 0.924	0.0914	0.796	0.685, 0.920	0.0912
100	$\alpha = 1/3$	$\lambda_A = 0.5$	0.805	0.624, 1.034	0.1529	0.804	0.624, 1.023	0.1516
		$\lambda_A = 2$	0.801	0.629, 1.027	0.1508	0.801	0.620, 1.024	0.1493
	$\alpha = 3$	$\lambda_A = 0.5$	0.794	0.612, 1.034	0.1550	0.794	0.692, 1.034	0.1539
		$\lambda_A = 2$	0.801	0.629, 1.047	0.1540	0.801	0.636, 1.040	0.1520
25	$\alpha = 1/3$	$\lambda_A = 0.5$	0.786	0.451, 1.345	0.3261	0.782	0.460, 1.334	0.3209
		$\lambda_A = 2$	0.799	0.483, 1.353	0.3127	0.800	0.494, 1.347	0.3079
	$\alpha = 3$	$\lambda_A = 0.5$	0.795	0.473, 1.363	0.3231	0.789	0.461, 1.349	0.3205
		$\lambda_A = 2$	0.810	0.485, 1.366	0.3202	0.805	0.474, 1.363	0.3151

[a] Number per group.
[b] Hazard ratio.
[c] Standard error.

Percentiles and the event time ratio

The percentiles of a Weibull are easily derived,

$$e^{-\lambda t^{\alpha}} = p \Rightarrow t_p = \left[\frac{\log\left(\frac{1}{p}\right)}{\lambda}\right]^{\frac{1}{\alpha}}$$

where t_p denotes the time taken to reach the p^{th} percentile. The relative difference in the time to achieving the p^{th} percentile between treatments A and B is

$$\frac{t_{Ap}}{t_{Bp}} = \frac{\lambda_B^{\frac{1}{\alpha_B}}}{\lambda_A^{\frac{1}{\alpha_A}}}$$

which, under proportional hazards, simplifies to the acceleration factor or event time ratio,

$$\kappa = \left[\frac{\lambda_B}{\lambda_A}\right]^{\frac{1}{\alpha}} = \left[\frac{1}{\theta}\right]^{\frac{1}{\alpha}} \tag{9}$$

Again, based on the parameterization in Eq. (6), the log event time ratio for an individual with covariates $\underline{x_i}$ relative to an individual with covariates $\underline{x_j}$ is $\underline{\beta}'(\underline{x_i}-\underline{x_j})$. In the case of just two treatments, the log event time ratio is simply given by β.

Note that Eqs. (8) and (9) demonstrate that, under proportionality, parameters describing changes in the log event time ratio are simply a scalar multiple of those describing changes in the log hazard ratio. As event time and event rate ratios are therefore linked by the shape parameter, it follows that if the hazard ratio can be estimated in a Weibull analysis, then so can the event time ratio.

Assessing proportionality in a Weibull analysis

In the analysis of survival data, graphical methods are routinely employed to assess the extent to which proportionality holds [4]. These methods may also be supplemented by a simple test for proportionality [16]. If data follow a Weibull distribution, then a direct, model-based test of proportionality can easily be achieved by comparison of shape parameters. If a Weibull is fitted separately for each treatment group, the two shape parameters, σ_1 and σ_2, say, together with their variances, can be independently estimated and compared.

To test the hypothesis $H_0 : \hat{\sigma}_1/\hat{\sigma}_2 \neq 1$, then

$$\frac{\left[\log\left(\frac{\hat{\sigma}_1}{\hat{\sigma}_2}\right)\right]^2}{\left[\frac{\hat{V}(\hat{\sigma}_1)}{\hat{\sigma}_1{}^2} + \frac{\hat{V}(\hat{\sigma}_2)}{\hat{\sigma}_2{}^2}\right]}$$

can be compared to a $\chi_1{}^2$ distribution. If shape parameters are found to differ significantly, then the null hypothesis of proportionality is rejected. In practice it may be more sensible to

examine the confidence interval for the possible extent of nonproportionality rather than relying on a significance test. This is because relatively mild departures from proportionality, such as late divergence of survivor functions, have little impact on inferences, especially if interpretation is confined solely to the time period of observation. Thus, in the analysis of clinical trial data, even when there is some modest departure from proportionality, it may still be reasonable to conclude that the event rate and event time ratio estimates show, on average, treatment differences over the period of study follow-up.

Assessing treatment differences when proportionality does not hold

While some interpretation of treatment effect estimates may be possible in the presence of modest nonproportionality, some statisticians will rightly feel unease in drawing conclusions. This being the case, the Weibull allows the hazard ratio to be plotted as a function of time, via Eq. (7). From this description of the hazard ratio, it is possible to compare treatments in terms of the average or integrated hazard over some time interval $(0-T)$. The integrated hazard is given by $\lambda T^{\alpha-1}$ so that the ratio of average hazards is given by $(\lambda_A/\lambda_B)T^{\alpha_A-\alpha_B}$. If a Weibull model is again fitted to each treatment group separately, the variance-covariance matrices can again be used to derive the standard error of the log of this quantity:

$$\hat{SE}\ log\left[\frac{\lambda_A}{\lambda_B}\,T^{\,\alpha_A-\alpha_B}\right] \cong \sqrt{\sum_{r=A,B} V[\log(\hat{\lambda}_r)] + T^2V[(\hat{\alpha}_r] + 2T\mathrm{Cov}[\log(\hat{\lambda}_r), \hat{\alpha}_r]}$$

The ratio of average hazards may then be plotted, with confidence limits, against time in order to explore how the averaged hazard ratio evolves with follow-up.

Predicting data maturation

The Weibull has been used in the field of engineering to predict the proportion of future failures after having observed a failure process to a given point in time [17]. In the context of clinical trials, predicting how deaths are likely to accumulate over time is often important, especially in the many trials designed with prespecified, event-driven interim analyses. In such trials, it is of great interest to accurately predict the time course of emerging deaths so that the appropriate resources can be put into place and to forewarn that perhaps additional follow-up beyond that envisaged at the outset, or at the previous analysis, is required to achieve the desired level of data maturity.

This can either be achieved in aggregate, for each treatment group, via simple extrapolation of the estimated survivor function, $e^{-\hat{\lambda}_r t^{\hat{\alpha}}}$, $r = A, B$, or by a more complex, individual patient based analysis as follows:

1. Assume an analysis has been performed with a mean follow-up time F, at which time d patients have died and $c = n-d$ are censored.
2. Consider the individual i with covariates $\underline{x}_i$, censored at time F. The probability that this individual survives to time $F+S$ is

$$p(T > F+S/T > F) = \frac{e^{-\lambda_i(F+S)^{\alpha}}}{e^{-\lambda_i F^{\alpha}}}$$

so that

$$F + S = \left[-\frac{ln(1-U)}{\lambda_i} + F^{\alpha}\right]^{\frac{1}{\alpha}} \tag{10}$$

where $U \sim U(0,1)$.

3. Survival times for the c censored individuals can be predicted if c deviates are randomly sampled from a $U(0,1)$ distribution, and substituted into Eq. (10). If, for the i^{th} patient, predicted survival exceeds $F + S$, then the patient remains censored; otherwise the patient is predicted to have died in the interval $(F, F + S]$.
4. Repeating 3, say, 1000 times, and averaging over repeats, provides an estimate of the number of additional deaths expected in the interval $(F, F + S]$.

This approach allows individual patient covariates to be used in predicting survival time, and so overall data maturity at a time S following the earlier analysis at time F. If a given level of maturity is required at the next analysis, the amount of additional follow-up needed can be estimated by trial and error.

An example

Analysis by both Cox's regression and the Weibull model is illustrated in the following example [18]. Patients with early prostate cancer were randomized to one of two treatments, active (bicalutamide 150 mg) or placebo. The primary endpoint was progression-free survival. The analysis took place at a minimum of 2 years and a median of 3 years follow-up. All patients were followed to disease progression or death irrespective of withdrawal of randomized therapy or addition of other, systemic therapies. Patients who remained progression free, or who were lost to follow-up at some earlier point in time, were censored. In addition to randomized treatment, four important, prospectively identified prognostic factors were included as covariates: these were primary background therapy (surgery, radiotherapy, or observation); log prostate-specific antigen level at diagnosis; stage of disease (either localized or locally advanced); and the degree of differentiation of disease (well, moderate, or poorly differentiated). The effect of primary therapy was captured in terms of contrasts between surgery versus radiotherapy and surgery versus observation. Similarly, the effect of degree of differentiation was captured in terms of contrasts between well versus moderately differentiated and well versus poorly differentiated.

A total of 1798 and 1805 patients were randomized to active and placebo treatments, respectively. At the time of the analysis, 181 and 293 events had accrued on active and placebo, respectively.

The results of the analysis are presented in Tables 2 and 3. Fig. 2 shows the Kaplan-Meier curves for the treatment groups, together with the fitted survivor function estimates from the Weibull analysis.

In Table 2, it is immediately obvious that both analyses provide very similar results, this being expected as discussed above. As the Weibull models log time, the parameter estimates, $\hat{\beta}$, represent event time ratios on the log scale. For example, patients with poorly differentiated disease were associated with a reduction in event time of approximately 33% (since $e^{-0.3973} = 0.67$) relative to those with well differentiated disease. As indicated in Eq. (8), division of $\hat{\beta}$ by $-\hat{\sigma}$ converts Weibull parameters from log event time ratios to log hazard ratios. Informal comparison of $-\hat{\beta}/\hat{\sigma}$ and $\hat{\gamma}$ indicates a close match between the Cox and Weibull models. This is as expected given that the Weibull distribution provides a close fit to the data.

Table 3 provides estimates of the treatment effect, both in terms of a hazard ratio and an event time ratio. From these results it can be seen that treatment with bicalutamide 150 mg significantly reduces the risk of progression compared to placebo by approximately 43% and, in doing so, significantly increases the progression-free survival interval by approximately 50%.

In terms of predicting how events might accrue over time, application of Eq. (10) indicates expected maturities of 21%, 28%, and 35% with additional follow-up of 1, 2, and 3 years, respectively.

The affect of departures from the Weibull distribution

Concerns may arise when using Weibull-based analyses in that the data collected may not conform exactly to a Weibull distribution. Simple graphical checks can be used to assess the extent to which data have a Weibull distribution and residual diagnostics can be also examined to assess goodness of fit [2,4].

Nevertheless, concerns may still be present that without a close distributional match, inferences based on a Weibull analysis may be misleading. However, for modest departures

Table 2. Results of Weibull and Cox analyses

	Weibull: modeling event time ratio				Cox: modeling log hazard ratio		
	$\hat{\beta}$	SE[a]$\hat{\beta}$	t	$-\hat{\beta}/\hat{\sigma}$	$\hat{\gamma}$	SE$\hat{\gamma}$	t
Intercept, μ	8.977	0.158					
Shape, σ[b]	0.7275	0.0307					
Randomized treatment[b]	0.4022	0.0706	5.70	−0.5529	−0.5544	0.0947	−5.85
Log PSA[c] at diagnosis	−0.2005	0.0352	−5.70	0.2756	0.2772	0.0471	5.89
Disease stage	0.3802	0.0746	5.10	−0.5226	−0.5265	0.1002	−5.25
Radiotherapy	−0.3184	0.0987	−3.23	0.4377	0.4382	0.1347	3.25
Observation	−0.6184	0.0837	−7.39	0.8500	0.8548	0.1096	7.80
Moderately differentiated	−0.1456	0.0891	−1.63	0.2001	0.1977	0.1222	1.62
Poorly differentiated	−0.3973	0.0937	−4.24	0.5461	0.5500	0.1275	4.31

[a] Standard error.

[b] Covariance between scale and treatment parameters was estimated to be 0.00047305.

[c] Prostate-specific antigen.

Table 3. Estimated hazard (HR) and event time ratios (ETR) for active relative to placebo

Cox			Weibull					
HR	SE[a]	95% CI[b]	HR	SE	95% CI	ETR	SE	95% CI
0.574	0.0947	0.477, 0.692	0.575	0.0947	0.477, 0.693	1.495	0.0706	1.302, 1.717

[a] Standard error.
[b] Confidence interval.

from a true Weibull, such concerns may largely be unwarranted, especially for hazard ratio estimation under proportionality.

To investigate hazard ratio estimates achieved via Weibull and Cox analyses irrespective of the true distribution for survival times, the following simulation approach was used.

Clinical trial data were simulated from lognormal, gamma, and piecewise exponential distributions. In each case, two treatments, *A* and *B*, say, were assumed and a random amount of censoring (10%) was incorporated. Parameters for each distribution were chosen so that the mean on treatment *A* was 6 months, say, and also so that variance of the lognormal and gamma distributions coincided. For the piecewise exponential, both treatments were assumed to have a common event rate for the first 3 months, diverging thereafter. Treatment differences, in terms of ratio of means, of 1.25 and 1.50 were used. To further reflect the clinical trial situation, uniform patient accrual over a 6-month period was simulated and a data cutoff

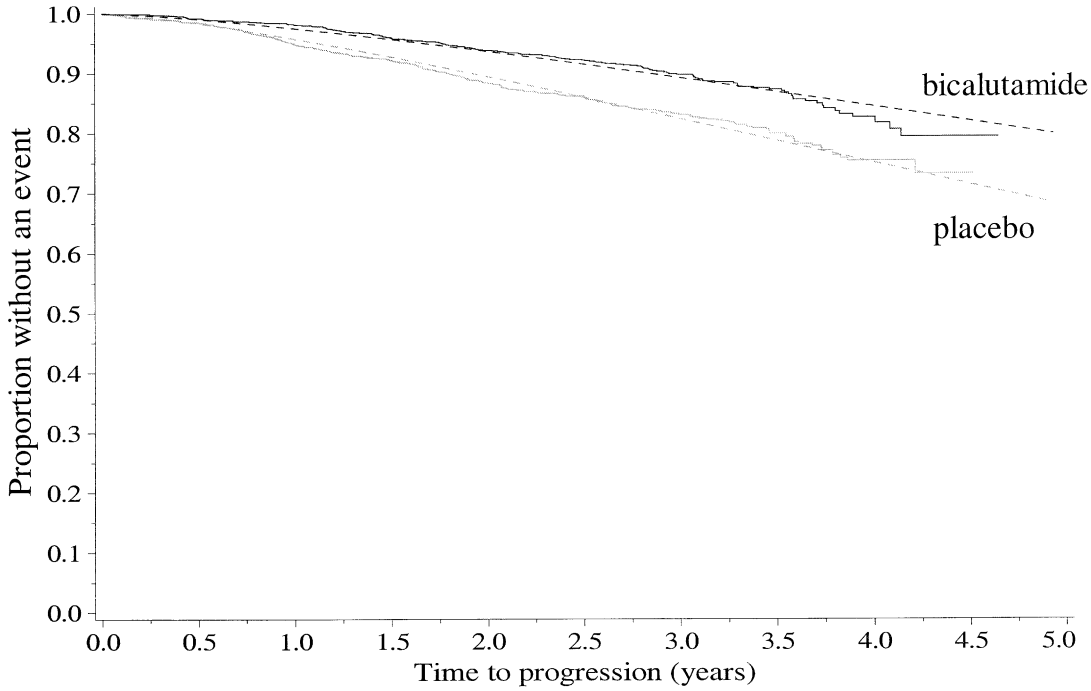

Fig. 2. Kaplan-Meier plot for progression-free survival with fitted Weibull survivor function estimates.

was employed whereby all event times were truncated at a given point in follow-up; the data cut-off time used was 12 months.

The lognormal and gamma survivor functions are illustrated in Figs. 3(a) and 3(b). Survivor function for the piecewise exponential distribution is illustrated in Fig. 3(c).

Data were then analyzed assuming a Weibull distribution via PROC LIFEREG in SAS. Both the hazard ratio and event time ratio were estimated. The data were also analyzed via PROC PHREG again in SAS to estimate the conventional Cox's hazard ratio. One thousand trial datasets, each of size $n = 400$ (200 in each treatment group), were simulated and the mean and variance of resulting log hazard and log event time ratios were calculated. A summary of results is provided in Tables 4–6.

These data suggest that even when data are known not to follow a Weibull distribution, analysis assuming a Weibull distribution provides results that are similar to those obtained by conventional Cox analysis. For the piecewise exponential, the Weibull analysis tends to give slightly larger log hazard ratio estimates (in absolute terms) compared to the Cox analysis, although the standard errors also tends to be slightly larger. Student t values were thus little different, perhaps being slightly higher for the Weibull analysis. With respect to the estimated event time ratio, this tends to be a little less than the known ratio of median times to event in the cases where the true median exceeds 3 months. If the 12-month data cutoff truncation is removed, then the resulting event time ratios are higher, as would be expected, and more in line with the known ratio of median times to event. For both lognormal and gamma data, results from Weibull and Cox analyses are virtually indistinguishable. The estimated event time ratio again tends to be less than the known ratio of median (or mean) times to event. Removal of the 12-month data cutoff truncation results in estimated event time ratios of 1.25 and 1.5, in exact concordance with the known differences in times to event.

This study suggests that hazard ratio estimates obtained via a Weibull analysis will tend to be similar to that obtained from a conventional Cox analysis, even when the Weibull does not provide an exact distributional match to the data. The importance of this is that for those data where it is considered reasonable to apply Cox regression to estimate the underlying hazard ratio, it should also be reasonable to apply a Weibull analysis to estimate the hazard ratio and, using the estimated scale parameter, to transform the hazard ratio to provide an estimated event time ratio. If the Weibull is considered only to provide a moderate fit to the data, then both the hazard and event time ratios can still be interpreted as the averaged risk of death and averaged increase in time on treatment A relative to treatment B. However, the estimation of percentiles, as described above, and the scale parameter are more dependent upon an adequate fit to the data. If a reasonable fit to the Weibull cannot be achieved, then it is recommended that percentiles be estimated directly from the Kaplan-Meier curves.

Summary

This paper has shown the Weibull model can provide a useful, parametric alternative to conventional Cox's regression modeling in the analysis of survival data. In addition to the hazard ratio, Weibull analysis provides a means of directly estimating the relative

improvement in survival time, the event time ratio. This quantification of treatment effect is of some clinical relevance and is likely to be better understood by some nonstatisticians than the conventional hazard ratio. Further, it has been shown that when data follow a Weibull distribution, Weibull analysis is asymptotically as efficient as Cox regression; both approaches give rise to similar hazard ratio estimates with the same standard error. Even when data are known not to follow a Weibull distribution, analysis assuming a Weibull distribution can

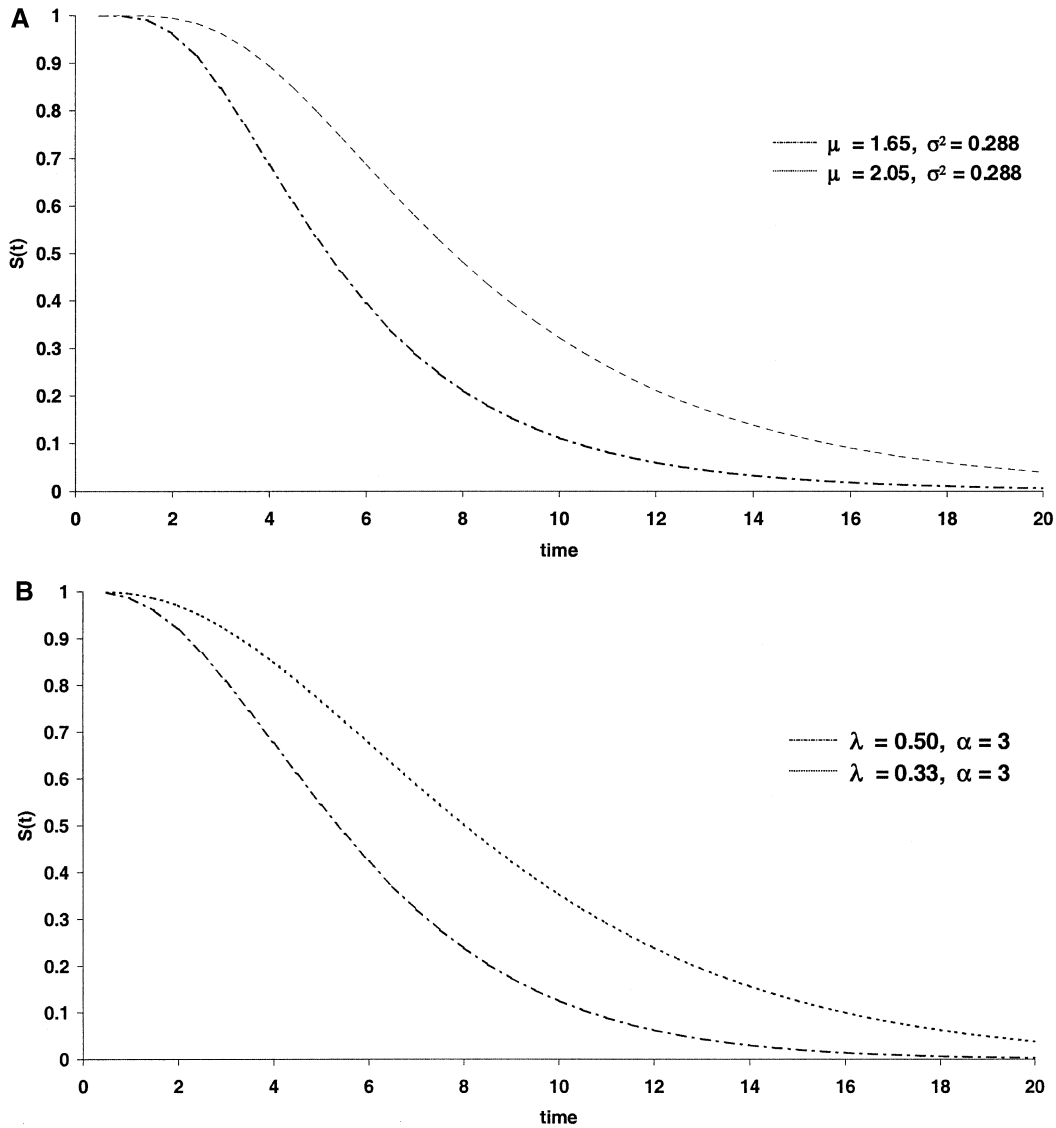

Fig. 3. (A) Lognormal survivor function. (B) Gamma survivor function.

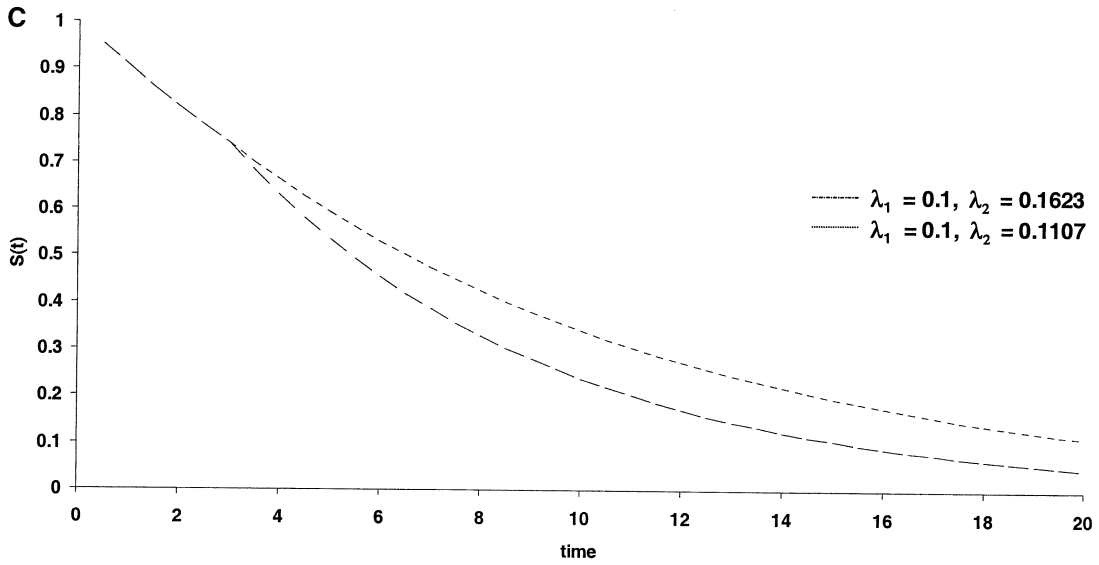

Fig. 3. (C) Piecewise exponential survivor function.

give very similar results, in terms of the hazard ratio, to those obtained by conventional Cox regression.

Key results in relation to the Weibull presented in this paper are thus:

1. Weibull analysis allows simultaneous characterization of the treatment effect in terms of the hazard ratio and the event time ratio, being a direct measure of the relative improvement in survival time.

Table 4. Simulation of piecewise exponential: analysis by Cox and by Weibull

			Cox analysis			Weibull analysis					
				SE^e			SE			SE	
$\lambda_1{}^a$	$\mu_A/\mu_B{}^b$	$\tilde{\mu}_A/\tilde{\mu}_B{}^c$	HR^d	*ln HR*	*t*	HR	ln HR	*t*	ETR^f	In ETR	*t*
0.01	1.25	1.13	0.834	0.1199	−1.51	0.826	0.1185	−1.62	1.099	0.0585	1.62
0.01	1.50	1.26	0.716	0.1270	−2.64	0.702	0.1251	−2.83	1.191	0.0612	2.86
0.10	1.25	1.10	0.872	0.1142	−1.19	0.874	0.1073	−1.25	1.115	0.0868	1.25
0.10	1.50	1.21	0.783	0.1195	−2.05	0.784	0.1123	−2.16	1.221	0.0920	2.17
1	1.25	1.00	0.995	0.1096	−0.05	0.982	0.1341	−0.33	1.033	0.1568	0.14
1	1.50	1.00	0.987	0.1127	−0.12	0.967	0.1407	−0.25	1.043	0.1658	0.25

[a] Common event rate over first 3 months.
[b] Ratio of mean times to event; μ_A = *6 months throughout.*
[c] Ratio of median times to event.
[d] Hazard ratio.
[e] Standard error.
[f] Event time ratio.

Table 5. Simulation of longnormal: analysis by Cox and by Weibull

		Cox analysis			Weibull analysis					
μ_A/μ_B[a]	σ^2	HR[b]	SE[c] ln HR	t	HR	SE ln HR	t	ETR[d]	SE ln ETR	t
1.25	0.288	0.676	0.1156	−3.39	0.662	0.1236	−3.33	1.209	0.0559	3.40
1.5	0.288	0.489	0.1158	−6.23	0.472	0.1203	−6.24	1.397	0.0515	6.49
1.25	1.386	0.833	0.1111	−1.64	0.827	0.1165	−1.64	1.199	0.1104	1.64
1.5	1.386	0.720	0.1128	−2.91	0.711	0.1154	−2.95	1.376	0.1075	2.97

[a] Ratio of mean times to event = ratio of median time-to-event for lognormal data. $\mu_A = 6$ months throughout.
[b] Hazard ratio.
[c] Standard error.
[d] Event time ratio.

2. Weibull and Cox analyses coincide when data follow a Weibull distribution; both approaches are asymptotically equally efficient.
3. The Weibull provides an adequate fit in many situations. Even when data do not follow an exact Weibull distribution, a Weibull-based analysis can give results that are very similar to those obtained from a Cox analysis. However, the estimation of percentiles and the event time ratio is more dependent upon an adequate fit to the data. If a reasonable fit to the Weibull cannot be achieved, then it is recommended that percentiles be estimated directly from the Kaplan-Meier curves.
4. Weibull analysis allows direct assessment and quantification of proportionality, or lack thereof.
5. If the data display nonproportional hazards, then a Weibull analysis provides a description of the hazard ratio (and event time ratio) over time and, depending on the circumstances, an analysis of hazards averaged over time.
6. Weibull analysis offers the opportunity to predict how data might mature over time, something that is of great interest within oncology trials, especially where a series of interim analyses are planned.

Table 6. Simulation of gamma: analysis by Cox and by Weibull

		Cox analysis			Weibull analysis					
μ_A/μ_B[a]	α	HR[b]	SE[c] ln HR	t	HR	SE ln HR	t	ETR[d]	SE ln ETR	t
1.25	3	0.679	0.1075	−3.60	0.675	0.1084	−3.62	1.212	0.0540	3.63
1.5	3	0.494	0.1170	−6.03	0.490	0.1141	−6.24	1.415	0.0539	6.45
1.25	1/3	0.897	0.1125	−0.96	0.901	0.1038	−1.00	1.258	0.2302	1.00
1.5	1/3	0.821	0.1189	−1.65	0.828	0.1103	−1.72	1.522	0.2441	1.72

[a] Ratio of mean times to event = ratio of median time-to-event for gamma data. $\mu_A = 6$ months throughout.
[b] Hazard ratio.
[c] Standard error.
[d] Event time ratio.

Discussion

The key implication of this paper is that in those very frequent instances where two or more treatments are to be compared for survival (or some other time-to-event endpoint) with adjustment for one or more baseline prognostic factors, the Weibull is at least as informative as a corresponding Cox analysis, and probably more so. Use of the Weibull provides researchers and data analysts with an estimate of treatment effect as per routine Cox analysis but, furthermore, provides a clinically useful, alternative representation of the treatment difference in terms of the event time ratio—consistency, in terms of statistical significance, is assured as both measures of treatment effect have essentially the same *p*-value. On this basis, and as the primary objective of clinical trials with survival as the primary endpoint is the simple comparison of survival distributions, it seems reasonable to argue that a Weibull-based analysis would likely serve data analysts and clinical researchers better than a corresponding Cox-based analysis in most circumstances.

The assumption of proportionality is often an issue that rightly concerns statisticians when analyzing via Cox, being explored heuristically via graphical methods. The use of the Weibull offers the ability to explicitly examine the degree and nature of proportionality and, further, allows a simple, direct test for its presence. These model utilities are likely to be valuable tools in the routine analysis of clinical trial data.

The quantification of the treatment effect along the time axis is, in the author's experience, one of the most common requests from clinicians and other nonstatisticians in the analysis of survival and other time-to-event data and, thus, is one of the most common disappointments with Cox-based analyses. Simultaneous estimation of effects in terms of both rate and time is therefore a key strength of Weibull-based analyses. Unless data follow an exponential distribution, the common use of the reciprocal of the hazard ratio as an estimate of the relative difference in the median times to event is incorrect as easily evidenced via the above example discussed previously where the reciprocal of the hazard ratio would suggest a 1.7-fold increase in progression-free survival time, whereas a more appropriate Weibull analysis gives an event time ratio of 1.5 [7]. The routine use of Kaplan-Meier curves as a descriptive aid to Cox or log-rank analyses is both standard and sensible, but all too often is taken by nonstatisticians to be the literal interpretation of the analysis, such that the *p*-value is "attached" to the curves rather than to the hazard ratio, a practice which can be misleading. Weibull analysis allows the survivor function to be estimated, which, when plotted, more accurately reflects the estimated treatment effect. This in turn allows prediction versus data maturation, something that is of considerable practical value in the ongoing management of clinical trials with time-to-event endpoints, and yet another feature that does not readily flow from conventional Cox-based regression.

In applying a Weibull analysis, concerns may arise regarding degree of model fit. The simulations carried out in this paper would suggest that the Weibull provides an adequate fit in many situations such that even when data do not follow an exact Weibull distribution, a Weibull-based analysis gives results similar to those obtained from a corresponding Cox analysis. Upon reflection, it is not surprising nor unexpected that a time-to-event model with the flexibility of both a shape (λ) and a scale parameter (α) would provide a good fit in many situations just as it is not surprising that the normal distribution, with both location (μ)

and scale (σ) parameters, provides an adequate fit to interval data in a wide variety of applications.

While the majority of time-to-event analyses in clinical trials are univariate in nature, multivariate data often arises and a reasonable question is whether a Weibull-based approach can offer advantages here, too. Extensions of Cox-based regression for repeated event data have been developed, such as the commonly used Andersen and Gill model, which assumes the risk of a repeat event is unaffected by earlier events and follows the proportional hazards assumption [19]. A comprehensive overview of Cox-based models for multiple failure-time data has been offered by Wei and Glidden [20]. These authors note that while Cox-type regression has been widely used for multivariate failure-time data, it may not fit data well, and AFT models offer a useful alternative. They also note that AFT models can accommodate repeat events without natural ordering, being in contrast to the Andersen and Gill approach where natural ordering is assumed. Indeed, as the Weibull and other AFT models are simply log-linear models with error distributions reflective of time-to-event data, existing and well-developed theory in relation to generalized linear models and multivariate data analysis can be applied [12].

When wishing to explore the relationship between multiple events, random effects or "frailty" analyses can be considered. While extensions to Cox-based analyses in the form of time-dependant covariates are possible, Weibull and AFT models are preferred by some authors [20]. Keiding et al have suggested it would be advantageous to upgrade AFT approaches alongside conventional approaches for random effects survival analyses, emphasizing intuitive interpretation of the Weibull model [21].

In addition to extension to multivariate failure-time data, the Weibull and other parametric models have been found to be useful in other areas also, such as data monitoring and analysis of failure-time data when cure is possible [22]. Sposto has examined parametric cure models, concluding that they are at least as good as Cox-based approaches and are to be preferred when proportionality fails to hold, allowing simultaneous assessment of covariate effects on both the proportion cured and the failure rate among those not cured [23].

Despite the many appealing features of Weibull-based analyses, the author does advocate wholesale replacement of Cox's proportional hazards regression model for routine, univariate failure-time analyses. Albeit at the cost of assuming proportionality, Cox's regression offers the advantage of being distribution-free and can readily accommodate time-dependent covariate analysis. Rather, a Weibull analysis offers the statistician the opportunity to supplement and enrich routine Cox regression analyses, especially when a direct quantification of improvement in survival time is desired or a more thorough evaluation of proportionality is warranted. Indeed, when such matters are of primary interest, one may reasonably argue that a Weibull-based approach is to be preferred, being at least as informative as Cox regression with no loss of power or sensitivity under proportional hazards. Therefore, the use of the Weibull, or other parametric models, in the analysis of survival data in clinical trials, at very least, should not be overlooked and even be promoted to sit with equal status alongside routine Cox-based analyses.

It is interesting to note that model-based analyses are the norm and have been for many decades, in the analysis of normally distributed data, where analysis of variance to multivariate analysis of covariance to complex nonlinear mixed effects modeling approaches are routinely

employed, with nonparametric alternatives mainly taking a supporting role. Similar comments can be made in relation to binary and ordered categorical data where model-based data analysis prevails. For survival and other time-to-event data, however, the approach presently taken is the inverse in many ways; nonparametric analyses are considered standard while potentially more informative model-based approaches are seldom seen. This may, in part, be due to past difficulties in computationally applying these models.

However, widely available software packages, such as SAS and S-Plus, have simple procedures devoted to data analysis via the Weibull and other parametric models such as the gamma and lognormal [24]. Application of the Weibull-based analyses described in this paper is therefore very straightforward and not an area where specific, homegrown software has to be written to affect an analysis. Hence, it is fair to say that statisticians have simple and readily accessible software on their desks and are thus well poised and better equipped than ever before to reap the benefits that Weibull and other parametric-based approaches have to offer in the day-to-day, practical analysis of survival data arising in clinical trials.

Acknowledgments

The author would like to thank Mr. Stuart Ellis, AstraZeneca Pharmaceuticals, Alderley Park, UK, and two anonymous referees for their thoughtful comments that greatly helped to improve the clarity and focus of this manuscript.

Appendix: The hazard ratio for a parametric failure-time distribution under proportional hazards

Let T denote the time-to-event variable with probability distribution function $f(t)$, and survivor function $S(t)$. Under proportional hazards

$$h_x(t) = h(t)e^{\alpha+x\gamma}$$

where $x = 1$ for treatment group A, $x = 0$ for treatment group B.

Assuming N_A patients in group A with d_A events and $N_A - d_A$ censored. Employing similar notation for group B, the likelihood for the data observed is

$$L = \prod_i^{d_A} h(t_i)e^{d_A(\alpha+\gamma)}\, e^{-\sum^{N_A}\int_0^{t_i} h(u)e^{\alpha+\gamma}du} \prod_i^{d_B} h(t_i)e^{d_B\alpha}\, e^{-\sum^{N_B}\int_0^{t_i} h(u)e^{\alpha}du}$$

Thus,

$$\ell = \log(L) \propto d_A(\alpha+\gamma) - \sum^{N_A}\int_0^{t_i} h(u)e^{\alpha+\gamma}du + d_B\alpha - \sum^{N_B}\int_0^{t_i} h(u)e^{\alpha}du$$

$$\frac{\partial \ell}{\partial \gamma} = d_A - e^{\alpha+\gamma}\sum^{N_A}\int_0^{t_i} h(u)du \tag{11}$$

$$\frac{\partial \ell}{\partial \alpha} = d_A - e^{\alpha+\gamma} \sum^{N_A} \int_0^{t_i} h(u)du + d_B - e^{\alpha} \sum^{N_B} \int_0^{t_i} h(u)du \tag{12}$$

$$\frac{\partial^2 \ell}{\partial \gamma \partial \alpha} = -e^{\alpha+\gamma} \sum^{N_A} \int_0^{t_i} h(u)du \tag{13}$$

$$\frac{\partial^2 \ell}{\partial \gamma^2} = -e^{\alpha+\gamma} \sum^{N_A} \int_0^{t_i} h(u)du \tag{14}$$

$$\frac{\partial^2 \ell}{\partial \alpha^2} = -e^{\alpha+\gamma} \sum^{N_A} \int_0^{t_i} h(u)du - e^{\alpha} \sum^{N_B} \int_0^{t_i} h(u)du \tag{15}$$

Hence, Eqs. (11) and (12) give

$$e^{\hat{\alpha}} = \frac{d_B}{\sum^{N_B} \int_0^{t_i} h(u)du}$$

$$e^{\hat{\alpha}+\hat{\gamma}} = \frac{d_A}{\sum^{N_A} \int_0^{t_i} h(u)du}$$

so that the hazard ratio, for any probability density function, $f(t)$, under the assumption of proportionality, is given by

$$e^{\hat{\gamma}} = \frac{\sum^{N_B} \int_0^{t_i} h(u)du}{\sum^{N_A} \int_0^{t_i} h(u)du} \frac{d_A}{d_B}$$

and Eqs. (13), (14), and (15) give

$$\underline{V} = E\left[-\underline{I}^{-1}\right]$$

$$-\underline{I}^{-1} = \begin{bmatrix} X+Y & Y \\ Y & Y \end{bmatrix}^{-1} = \frac{1}{XY} \begin{bmatrix} Y & -Y \\ -Y & X+Y \end{bmatrix}$$

where

$$X = e^{\alpha}\sum^{N_B}\int_0^{t_i} h(u)du \quad \text{and} \quad Y = e^{\alpha+\gamma}\sum^{N_A}\int_0^{t_i} h(u)du$$

Therefore, the variance of the log hazard ratio, under the assumption of proportionality, is given by

$$\hat{V}(\hat{\gamma}) = E\left[\frac{1}{X} + \frac{1}{Y}\right] = \left[\frac{1}{e^{\hat{\alpha}}\sum^{N_B}\int_0^{t_i} h(u)du}\right] + \left[\frac{1}{e^{\hat{\alpha}+\hat{\gamma}}\sum^{N_A}\int_0^{t_i} h(u)du}\right] = \frac{1}{d_A} + \frac{1}{d_B}$$

If γ is small, then

$\hat{V}(\hat{\gamma}) \cong 4/d$ where d denotes the total number of events across both groups.

References

[1] Cox DR. Regression models and life tables (with discussion). Journal of the Royal Statistical Society, Series B 1972;34:187–220.
[2] Kalbfleisch JD, Prentice RL. The statistical analysis of failure-time data. New York: Wiley, 1980.
[3] Cox DR, Oakes D. Analysis of survival data. London: Chapman and Hall, 1984.
[4] Collett D. Modeling survival data in medical research. London: Chapman and Hall, 1994.
[5] Peto R, Peto J. Asymptotically efficient rank invariant test procedures (with discussion). Journal of the Royal Statistical Society, Series A 1972;135:185–207.
[6] Armitage P, Berry G. Statistical methods in medical research. Oxford: Blackwell Scientific Publications, 1987.
[7] Sylvester R, Collette L. When do statistics lie? UroOncology 2001;1:185–194.
[8] Kim J. Confidence intervals for the difference of median survival times using the stratified Cox proportional hazards model. Biometrical Journal 2001;43:781–790.
[9] Prentice RL, Kalbfleisch JD. Hazard rate models with covariates. Biometrics 1979;35:25–39.
[10] Reid N. A conversation with Sir David Cox. Statistical Science 1994;9:439–455.
[11] Byar DP. Analysis of survival data: Cox and Weibull models with covariates. Statistics in Medical Research 1982;12:365–401.
[12] Wei LJ. The accelerated failure-time model: a useful alternative to the Cox regression model in survival analysis. Stat Med 1992;11:1871–1879.
[13] Chen YQ, Wang M-C. Analysis of accelerated hazards model. J Am Stat Assoc 2000;95:608–618.
[14] Chen YQ, Wang M-C. Estimating a treatment effect with the accelerated hazards model. Control Clin Trials 2000;21:369–380.
[15] SAS/STAT User's Guide. Version 6, 4th ed, Volume 2. Cary, NC: SAS Institute Inc., 1989.
[16] Gill RD, Schumacher M. A simple test of the proportional hazards assumption. Biometrika 1987;74:289–300.
[17] Nelson W. Weibull prediction of a future number of failures. Quality and Reliability Engineering International 2000;16:23–26.
[18] Wirth M, Tyrrell C, Wallace M, et al. Bicalutamide (Casodex) 150 mg as immediate therapy in patients with localized or locally advanced prostate cancer significantly reduces risk of disease progression (with Editorial comment). Urology 2001;58:146–151.
[19] Andersen PK, Gill RD. Cox's regression model for counting processes: a large sample study. Ann Stat 1982;10:1100–1120.

[20] Wei LJ, Glidden DV. An overview of statistical methods for multiple failure-time data in clinical trials. Stat Med 1997;16:833–839.
[21] Keiding N, Andersen PK, Klein JP. The role of frailty models and accelerated failure-time models in describing heterogeneity due to omitted covariates. Stat Med 1997;16:215–224.
[22] Lecoutre B, Mabika B, Derzko G. Assessment and monitoring in clinical trials when survival curves have distinct shapes. Stat Med 2002;21:663–674.
[23] Sposto R. Cure model analysis in cancer: an application to data from the Children's Cancer Group. Stat Med 2002;21:293–312.
[24] S-PLUS 6 Guide to Statistics, Vol. 2. Seattle, Washington: Insightful Corporation, 2001.