Lifetime Data Analysis, 10, 335–350, 2004

Regression Analysis of Restricted Mean Survival Time Based on Pseudo-Observations

PER KRAGH ANDERSEN pka@biostat.ku.dk
Department of Biostatistics, University of Copenhagen, Blegdamsvej 3, DK 2200 Copenhagen N, Denmark

METTE GERSTER HANSEN
Department of Biostatistics, University of Copenhagen,Blegdamsvej 3, DK 2200 Copenhagen N, Denmark

JOHN P. KLEIN
Division of Biostatistics, Medical College of Wisconsin, 9701 Watertown Plank Road, Milwaukee WI 53226, USA

Received December 2, 2003; Revised June 22, 2004; Accepted June 29, 2004

Abstract. Regression models for survival data are often specified from the hazard function while classical regression analysis of quantitative outcomes focuses on the mean value (possibly after suitable transformations). Methods for regression analysis of mean survival time and the related quantity, the restricted mean survival time, are reviewed and compared to a method based on pseudo-observations. Both Monte Carlo simulations and two real data sets are studied. It is concluded that while existing methods may be superior for analysis of the mean, pseudo-observations seem well suited when the restricted mean is studied.

Keywords: censoring, hazard function, health economics, mean survival time, pseudo-observations, regression model, restricted mean survival time, survival analysis

1. Introduction

Regression models for survival data are frequently specified via the hazard function, $\alpha(\cdot)$, for the distribution of the survival time X. This is given by

$$\alpha(t \mid \mathbf{Z}) = -\frac{d}{dt}\log S(t \mid \mathbf{Z}),$$

where $S(t \mid \mathbf{Z}) = \mathrm{pr}(X > t \mid \mathbf{Z})$ is the survival function given the covariates, $\mathbf{Z}$. Examples include the Cox (1972) proportional hazards model

$$\alpha(t \mid \mathbf{Z}) = \alpha_0(t)\exp(\beta^\top \mathbf{Z}) \tag{1}$$

and Aalen's (1980, 1989) non-parametric additive hazards model

$$\alpha(t \mid \mathbf{Z}) = \alpha_0(t) + \beta(t)^\top \mathbf{Z}, \tag{2}$$

where $\alpha_0(t)$ is an unspecified baseline hazard and $\beta(t)$ a vector of unspecified regression functions. A semi-parametric alternative to (2) is the Lin and Ying (1994) model with $\beta(t) = \beta$. In these models, effects of treatment and other covariates are hazard ratios or hazard differences, respectively.

In classical linear regression of quantitative outcomes, focus is on the mean value and covariate effects are differences between mean values. In some studies where survival time is the outcome variable it would be appealing to be able to express covariate effects on a mean survival time scale (alternatively, on a mean log-survival time scale). Examples include studies in health economics where, for example, the costs associated with patients being hospitalized are directly dependent on the average length of stay in hospital, e.g., Li (1999). The mean survival time $\mu = \mathrm{E}X$ (if it exists) is given by

$$\mu = \int_0^\infty S(t)dt. \tag{3}$$

However, because of the inevitable right-censoring present in studies of survival time, the tail of the survival time distribution, and thereby the mean survival time, may often be ill-determined. As an alternative to μ, it has been suggested to study the restricted mean survival time which for any $\tau > 0$ is given by

$$\mu(\tau) = \mathrm{E}(X \wedge \tau) = \int_0^\tau S(t)dt. \tag{4}$$

The first use of this functional seems to be by Irwin (1949).

Having specified a model $\alpha(t \mid \mathbf{Z})$ for the hazard, a model is also implied for the survival function $S(t \mid \mathbf{Z})$ and, via (3) and (4), for the mean survival time $\mu(\mathbf{Z})$ and for the restricted mean survival time $\mu(\tau, \mathbf{Z})$. However, if a simple regression model like (1) or (2) holds for the hazard then the way in which the mean $\mu(\mathbf{Z})$ or the restricted mean $\mu(\tau, \mathbf{Z})$ depends on the covariates is generally not described by simple parameters.

Much research has gone into direct regression analysis of mean (log-) survival time, one prominent class of models being the *accelerated failure time models*. This is an important special case of the *transformation model*

$$h(X_i) = \beta_0 + \beta^\top \mathbf{Z}_i + \varepsilon_i \tag{5}$$

(e.g., Dabrowska and Doksum, 1988; Fine et al., 1998), where $h = \log$. For $h = id$, the identity function, this is just a linear model for mean survival. In this representation, the residuals $\varepsilon_i, i = 1, \ldots, n$ are independent and identically distributed zero-mean random variables; i.e., for the accelerated failure time model,

$$\mathrm{E}(\log X_i) = \beta_0 + \beta^\top \mathbf{Z}_i.$$

This is just a standard linear model for $\log X_i$ but estimation procedures are needed to account for right-censoring. For a parametric specification of the residuals maximum likelihood estimation is straightforward and several computer packages have implemented methods when the residuals have an extreme value distribution (corresponding to the distribution of X_i being *Weibull*), or a log-normal or a log-logistic distribution. For the semiparametric case when the residual distribution is not specified, a generalized least squares procedure was developed by Buckley and James (1979) with asymptotic results by Lai and Ying (1991). An alternative estimation procedure based on rank tests (in fact, shown by Ritov, 1990, to be asymptotically equivalent to the Buckley–James method) was studied by, among others, Tsiatis (1990), Wei et al. (1990) and Ying (1993).

Another line of research initiated by Leurgans (1987) uses so-called "synthetic data" to analyse regression models for the mean survival time. In this approach, suppose that the true uncensored survival times are $X_i, i = 1, \ldots, n$, while the observed data are $(\widetilde{X}_i, D_i), i = 1, \ldots, n$, where $D_i = I(X_i = \widetilde{X}_i)$ are failure indicators and $X_i > \widetilde{X}_i$ when $D_i = 0$. The idea is then to replace the observed times of observations by a synthetic sample $X_i^*, i = 1, \ldots, n$, such that $\mathrm{E}X_i = \mathrm{E}X_i^*$ and on which regression analysis may be performed. This sample may be obtained as

$$X_{(i)}^* = \widetilde{X_{(1)}} + \int_{\widetilde{X_{(1)}}}^{\widetilde{X_{(i)}}} \{\widehat{H(s)}\}^{-1} ds, \tag{6}$$

where $\widehat{H(s)}$ is the Kaplan–Meier estimator for the censoring distribution and $\widetilde{X_{(1)}} \leq \cdots \leq \widetilde{X_{(n)}}$ are the ordered times of observation. This approach was further developed by Zhou (1992) and Zheng (1995). It is seen that observed, true failure times smaller than the smallest censored observation are left unchanged while larger times of observation (both censored and uncensored) are spread out compared to the observed data, the amount of spread depending on the number of smaller censored observations. An example is given in Section 2.

In contrast, relatively little work has been done concerning regression analysis of restricted mean survival. Karrison (1987) studied a proportional hazards model (1) with a piecewise constant baseline hazard and used the implied regression model for $\mu(\tau, \mathbf{Z})$ obtained from (4). His approach was generalized by Zucker (1998) who took the standard semi-parametric proportional hazards model as his starting point. However, as mentioned above, in these models the relationship between $\mu(\tau, \mathbf{Z})$ and a given covariate is not described by simple regression parameters. Finally, Chen and Tsiatis (2001) studied methods for comparing covariate-adjusted restricted mean survival times between two treatment groups.

The purpose of the present paper is to study a different approach to analysis of both mean and restricted mean survival, namely by using so-called *pseudo-observations*, see Andersen et al. (2003). We will investigate the performance of estimators based on pseudo-observations for both mean and restricted mean survival time using Monte Carlo simulations and we will compare with other methods for the mean. The use of pseudo-observations will also be illustrated using two real data sets.

The structure of the paper is as follows. In Section 2 we introduce pseudo-observations, specialize to mean and restricted mean survival, and present the examples. Section 3 contains the simulation study and some concluding remarks and further discussion are found in Section 4.

2. Pseudo-Observations

Following Andersen et al. (2003), pseudo-observations are defined in the following way. Let $X_i, i = 1, \ldots, n$, be independent and identically distributed random variables, let θ be a parameter of the form

$$\theta = \mathrm{E}f(X_i) \tag{7}$$

and assume that we have an (at least approximately) unbiased estimator $\widehat{\theta}$ for this parameter. Let, furthermore, $\mathbf{Z}_i, i = 1, \ldots, n$, be independent and identically distributed covariates and define the conditional expectation

$$\theta_i = \mathrm{E}\{f(X_i) \mid \mathbf{Z}_i\}.$$

The ith pseudo-observation is then

$$\widehat{\theta}_i = n \cdot \widehat{\theta} - (n-1) \cdot \widehat{\theta^{-i}}, \tag{8}$$

where $\widehat{\theta^{-i}}$ is the "leave-one-out" estimator for θ based on $X_j, j \neq i$. Note that if all X_i are observed then θ may be estimated by the average of the $f(X_i)$ in which case $\widehat{\theta}_i$ is simply $f(X_i)$. We shall be using this approach in a situation where only a censored sample of the X_i is available.

A regression model for the parameter θ corresponds to a specification of how θ_i depends on $\mathbf{Z}_i$ and this may done via a generalized linear model

$$g(\theta_i) = \beta^T \mathbf{Z}_i \tag{9}$$

with link function $g(\cdot)$ (where a column $Z_{i0} = 1$ has been added to $\mathbf{Z}_i$ corresponding to an intercept, β_0). The regression coefficients, β, may now be estimated from the *generalized estimating equations*

$$\begin{aligned} U(\beta) &= \sum_{i=1}^{n} U_i(\beta) \\ &= \sum_{i=1}^{n} \left(\frac{\partial}{\partial \beta} g^{-1}(\beta^{\mathrm{T}} \mathbf{Z}_i)) V_i^{-1} (\widehat{\theta}_i - g^{-1}(\beta^{\mathrm{T}} \mathbf{Z}_i) \right) \\ &= 0. \end{aligned} \tag{10}$$

In the general situation, θ may be multivariate in which case V_i is a working covariance matrix for $\widehat{\theta}_i$ (Liang and Zeger, 1986; Zeger and Liang, 1986). In our present application, θ is a scalar and V_i is just a working variance of $\widehat{\theta}_i$.

Andersen et al. (2003) showed that consistent estimates of β could be obtained from (10) and that variance estimates for the solution $\widehat{\beta}$ could be obtained from the standard sandwich estimator

$$\hat{\Sigma} = I(\widehat{\beta})^{-1} \hat{\mathrm{var}}\{U(\beta)\} I(\widehat{\beta})^{-1}, \tag{11}$$

where

$$I(\beta) = \sum_i \left(\frac{\partial g^{-1}(\beta^{\mathrm{T}} \mathbf{Z}_i)}{\partial \beta} \right)^{\top} V_i^{-1} \left(\frac{\partial g^{-1}(\beta^{\mathrm{T}} \mathbf{Z}_i)}{\partial \beta} \right),$$

$$\hat{\mathrm{var}}\{U(\beta)\} = \sum_i U_i(\widehat{\beta}) U_i(\widehat{\beta})^{\mathrm{T}}.$$

This means that once the pseudo-observations are computed, estimation of β and Σ may be carried out using standard statistical software like SAS's PROC GENMOD.

The purpose of the present paper is to use this approach for analysis of the mean survival time $\mu(\mathbf{Z})$ and the restricted mean survival time $\mu(\tau, \mathbf{Z})$ defined in (3) and (4). Let $\widehat{S(t)}$ be the Kaplan–Meier (1958) estimator for the survival function $S(t)$ based on a right-censored sample $\widetilde{X}_i, i = 1, \ldots, n$, of the survival times $X_i, i = 1, \ldots, n$. Following Gill (1983), who studied the large-sample properties of estimates for μ obtained by plugging $\widehat{S(t)}$ into (3), we will use the version of $\widehat{S(t)}$ which is set to 0 for $t > \widetilde{X_{(n)}}$, the largest observation time. We will, therefore, base our inference for μ and $\mu(\tau)$ on the estimators obtained using this version of the Kaplan–Meier estimator

$$\widehat{\mu}(\tau) = \int_0^\tau \widehat{S(t)} \mathrm{d}t$$

with $\widehat{\mu} = \widehat{\mu}(\infty) = \widehat{\mu}(\widetilde{X_{(n)}})$. Pseudo-observations are defined accordingly from (8) and analysed using (10) and (11) with the simple choice, $V_i = 1$, of working variance.

We will now first, for illustration, show both the synthetic data (6) and the pseudo-observations (8) for the classical Freireich data (e.g., Cox, 1972) and next present two examples comparing the pseudo-observation method to other approaches.

Briefly, Freireich's data deal with a trial in childhood leukemia comparing length of remission in (paired) groups treated with 6-MP or placebo. In the placebo group,

Table 1. Remission times from the 6-MP treatment group of Freireich's data (+ denotes a censored observation) together with synthetic data (6) and pseudo-observations (8).

Observed Remission Length (weeks)	Synthetic Data	Pseudo-Observation
6,6,6	6	6
6+	6	26.17
7	7.1	5.8
9+	9.2	27.44
10	10.3	7.67
10+	10.3	28.85
11+	11.5	28.85
13	14.1	7.86
16	18.0	11.89
17+	19.3	32.65
19+	22.2	32.65
20+	23.8	32.65
22	27.5	13.62
23	29.4	15.64
25+	33.1	39.86
32+, 32+	49.3	39.86
34+	58.6	39.86
35+	67.9	39.86

no observations were censored. Table 1 shows the observed times in remission for the 6-MP treatment group together with the synthetic data and the pseudo-observations. It is seen that while the transformation into synthetic values are similar for censored and uncensored data the pseudo-observations are quite different for the two types of data. A censored time gives rise to a large pseudo-value while pseudo-observations for uncensored data are smaller than the actually observed value. Note also that the pseudo-observations for all censored observations between two successive failure times and larger than the largest failure time are identical. That the pseudo-observation for the largest censored observation, $\widetilde{X_{(n)}}$, is the same as for $\widetilde{X_{(n-1)}}$ is due to the fact that we always integrate to $\widetilde{X_{(n)}}$ when estimating the mean.

Example 1. The CSL1 Trial in Liver Cirrhosis

CSL1 was a double blind multicentre randomised clinical trial with the purpose of studying the effect of prednisone treatment versus placebo on survival in patients with liver cirrhosis. The accrual period ranged from 1962 to 1969 and the patients were followed to death, censoring or to September 1974, 292 patients were observed to die. Schlichting et al. (1983) and Christensen et al. (1985) analysed data from 488 patients recruited in this period of whom 102 had ascites (excess fluid in the abdomen). One of the early findings from this trial (CSL, 1974) was that patients without ascites seemed to benefit from the prednisone treatment while this treatment was harmful for patients with ascites. This is illustrated by fitting a Cox regression model to the data also adjusting for age (see Table 2).

For patients without ascites the estimated hazard ratio for treatment (prednisone vs. placebo) is 0.77 while that for patients with ascites is 1.61. The partial likelihood ratio test for no interaction is 8.96 (1 d.f., $P < 0.005$). In the same model the hazard ratio for ascites vs. no ascites in the placebo group is 1.45 and that for age is 1.04 per year.

Table 2. Estimated effects (with 95% confidence limits) for prednisone treatment, ascites and age based on the CSL1 data. W: "Weibull" regression model (5), P: pseudo-observations using estimating equations (10). For the Cox model estimates are hazard ratios, for the models for log X (middle panel) estimates are multiplicative effects on the life length, and for the models for X (lower panel) estimates are additive effects (in years) on the life length.

Method	Prednisone: No Ascites	Prednisone Ascites	Ascites Placebo	Age Per Year
Cox Model	0.77 (0.59, 1.01)	1.61 (1.08, 2.40)	1.45 (1.17, 1.80)	1.04 (1.03, 1.06)
W:E (log X)	1.32(0.97, 1.79)	0.57(0.37, 0.89)	0.66(0.52, 0.84)	0.95(0.94, 0.97)
P:E (log X)	1.25 (0.97, 1.61)	0.52 (0.27, 1.01)	0.53 (0.30, 0.92)	0.96 (0.95, 0.97)
P:E log ($X \wedge 5$)	1.17 (0.96, 1.43)	0.56 (0.31, 1.02)	0.58 (0.36, 0.95)	0.97 (0.96, 0.98)
W:E (X)	1.11 (0.18, 2.03)	−1.34 (−2.74, 0.07)	−1.25 (−2.00, −0.51)	−0.16 (−0.21, −0.12)
P:E (X)	0.76 (−0.32, 1.84)	−1.46 (−3.20, 0.28)	−1.41 (−3.13, 0.31)	−0.15 (−0.20, −0.11)
P:E ($X \wedge 5$)	0.38 (0.01, 0.75)	−0.83 (−1.61, −0.05)	−0.63 (−1.31, 0.05)	−0.05 (−0.07, −0.04)

It is of interest to see how the treatment effects for patients with or without ascites are when mean survival time (or mean log survival time) is the parameter of interest. To evaluate this we fitted parametric models of the form (5) with ε_i extreme value distributed and with $h =$ identity and $h = \log$ using SAS PROC LIFEREG. When $h = \log$ these are Weibull regression models. The resulting estimates are found in Table 2. It is seen that, in the model for mean log survival, prednisone treatment increases survival by a factor of 1.32 $(= \exp(\widehat{\beta}))$ for patients without ascites and by a factor 0.57 for patients with ascites. For the model for mean survival the treatment effect is $\widehat{\beta} = +1.11$ years (0.18, 2.03) and -1.34 years (-2.74,0.07), respectively, in the two groups. The latter estimates and their confidence limits may be directly compared with the results obtained using pseudo-observations where we find $+0.76$ years (-0.32, 1.84) and -1.46 years (-3.20, 0.28) (Table 2). To analyse $\mathrm{E}(\log(X))$ using pseudo-observations an approximately unbiased estimator for this parameter based on the whole sample is needed. Using integration by parts it is seen that this may be obtained from the integrated Kaplan–Meier estimator for $\log X$ as

$$\log(\widetilde{X_{(1)}}) + \int_{\log(\widetilde{X_{(1)}})}^{\log(\widetilde{X_{(n)}})} \widehat{S}(l)dl. \tag{12}$$

The treatment effects were quite close to those based on the Weibull distribution: 1.25 (0.97, 1.61) and 0.52 (0.27, 1.01), respectively (Table 2).

All the results quoted so far make assumptions about the mean of the survival time distribution, either that the distribution is Weibull or that $\widehat{\mu}(\widetilde{X_{(n)}})$ provides a sensible estimate. However, since the estimated survival probability at the largest observation time is 0.19 at 13.4 years in the placebo group and 0.15 at 12.2 years in the prednisone group one may argue that too little information is available to analyse the mean. Therefore, also analyses of the restricted mean at $\tau = 5$ years are presented (Table 2) both for X and for $\log X$. For X we find that prednisone treated patients without ascites gain 0.38 years (0.01, 0.75) during the first 5 years of treatment compared to the placebo group while prednisone treated patients with ascites lose 0.83 years (-1.61, -0.05). The same pattern is seen when analysing $\log X$. Provided that an additive model for the restriced mean is reasonable these results may be more reliable than those for the mean. To evaluate the model for the restricted mean the pseudo-observations were plotted against age in each of the four treatment by ascites groups with a lowess scatterplot smoother superimposed (see Figure 1).

The smooth curves should be approximately parallel straight lines which seems to be a reasonable approximation. Similar curves of average pseudo-values in three age groups (Figure 2) also suggest a satisfactory fit of the model.

In conclusion, using pseudo-observations we have been able to re-analyze the mean survival time in the CSL1 trial and compare with classical parametric models and, furthermore, we have been able to study regression models for the restricted mean survival time.

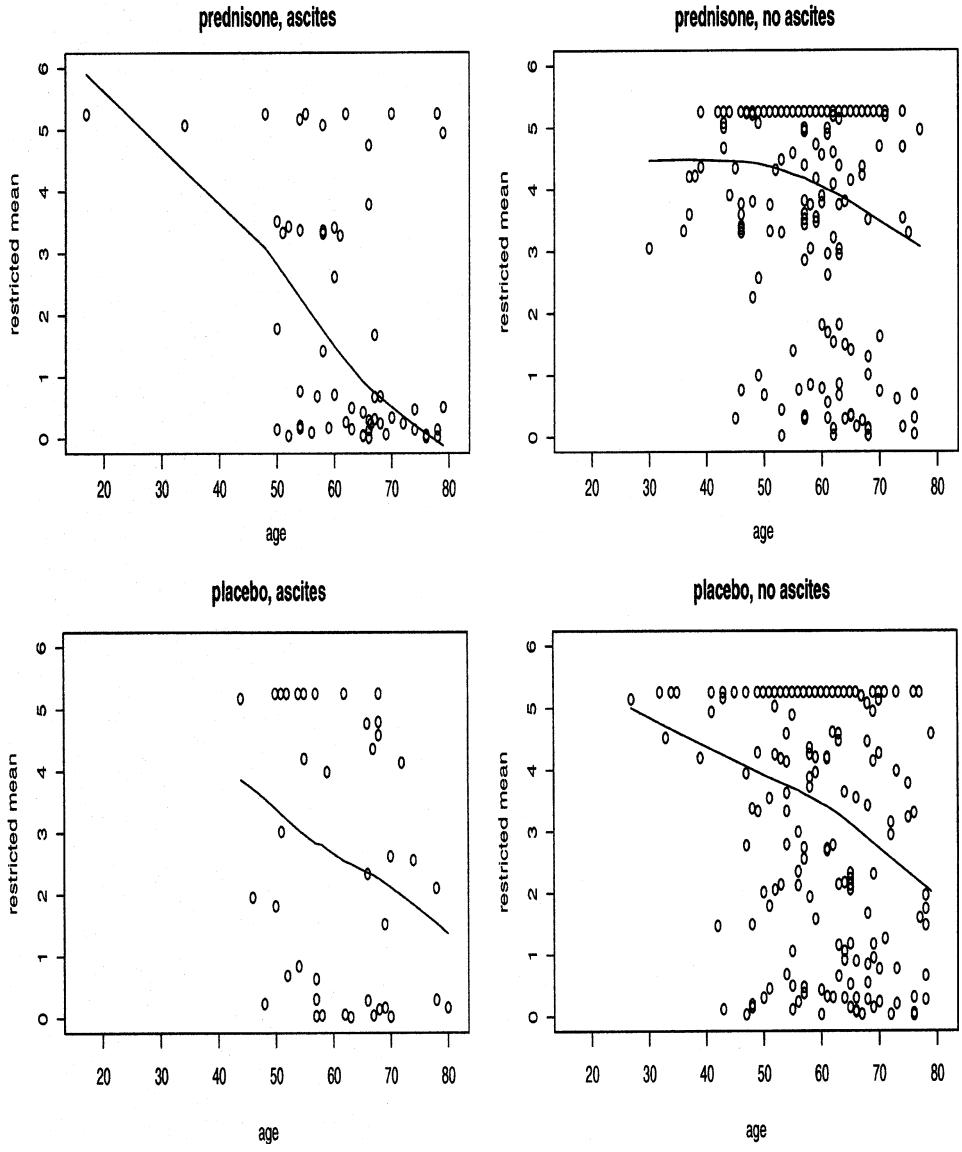

Figure 1. Pseudo-observations for the restricted mean life time plotted against age with a `lowess` smoother, CSL1 data.

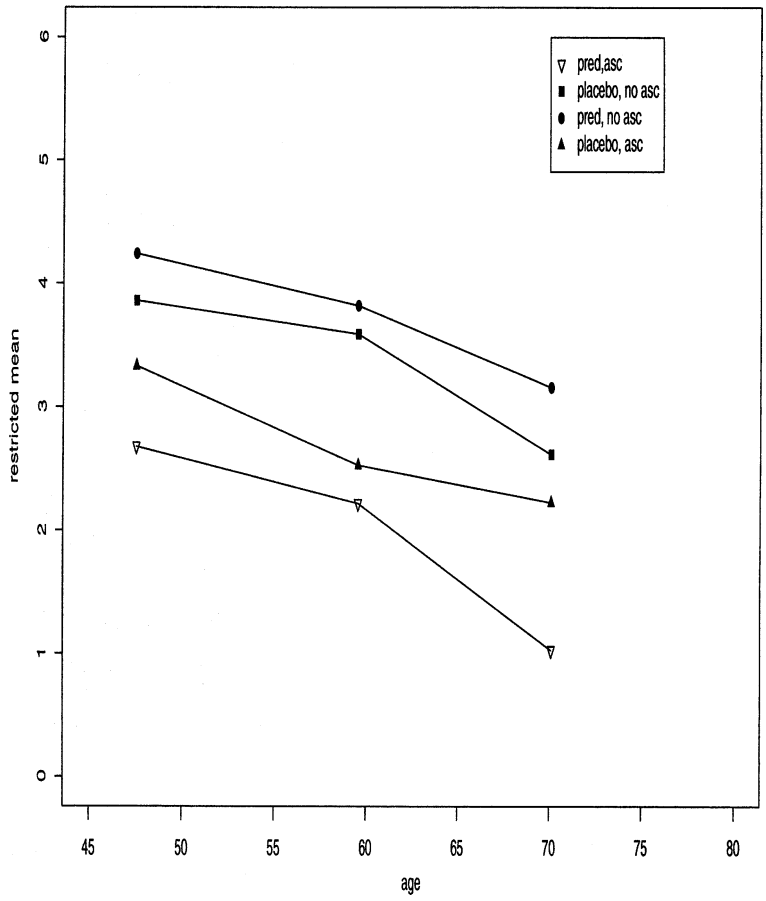

Figure 2. Average values of pseudo-observations for the restricted mean life time in three age groups plotted against age, CSL1 data.

Example 2. The Department of Biostatistics Dart League

In the mid-1990s a game of dart was played every day at lunch time at the Department of Biostatistics, University of Copenhagen. It varied from day to day which players took part in the game. On each day it was recorded who played that day, who finished the game, and how many rounds it took to finish that day's game. Thus, every player "runs his own race" but is stopped when the game is finished and, therefore, the data for each player (followed over time) consisted of the number of rounds played each day and an indicator of whether he or she finished the game that day. These data can be treated as a right-censored sample of the time (number of rounds) until the player finished the day's game and they can be summarized as a

survival curve. Since all games were truncated at 12 rounds the area under the survival curve up to $\tau = 12$ rounds was taken as a measure of how well the player performed; this number is the estimated restricted mean up to 12 rounds. Several factors could influence the number of rounds a player needs to finish the game on a given day and thereby the score. These include the number of players playing that day (a large number of players could sharpen competition and reduce the number of rounds), the day of the week, and the player's "handicap" on that day. To equalize the playing field, players with many victories were given one of two levels of handicap to make it more difficult for them to win.

Here we analyze the data from one player (the first author of the present paper, PKA) who played 212 times in 1994–1996 and finished in 82 games giving a score of 9.042. We shall examine the effect of the number of players on a day (≤ 5, 6–7, ≥ 8) and the level of handicap (none, first level, second level) on PKA's ability to finish the game. The simplest model to use would be a discrete-time Cox regression model with a complementary log–log link. However, the Cox model is modelling the rate at which a player is finishing and not the parameter of interest, the player's score. Regression models for this parameter are easier to interpret since they tell us how many additional or fewer rounds the player needs to finish. We focus on the score of PKA and compute pseudo-values for the mean time to finishing truncated at 12 rounds. Due to the discrete nature of the data there are few different values of the pseudo-observations corresponding to a finish or a censoring after 4, 5, ... ,12 rounds. These values are depicted in Figure 3. Like for the Freireich data of Table 1 we see that pseudo-observations for uncensored data points are somewhat smaller than the observed times whereas those corresponding to censored observations are substantially larger.

Applying the linear model (9) we find that the estimates for "handicap" were first vs. none 0.80 (95% c.i. −0.04 to 1.63), second vs. none 1.13 (95% c.i. −0.08 to 2.34), while the effect of "number of players" was quite small and insignificant. The two degree of freedom Wald type test of no effect of handicap using (11) yields a chi-square value of 5.14 ($P = 0.08$) while a 1 d.f. Wald type test for trend yields a chi-square value of 4.86 ($P = 0.03$). The major advantage of this approach is that the regression coefficients are directly interpretable as the number of extra rounds (up to 12) needed by PKA to finish the game with each of the two levels of handicap. The corresponding hazard ratio estimates from the discrete time Cox model would have a less direct interpretation.

3. Simulation Studies

3.1. Mean Survival Time

To study the behavior of estimators for mean survival time a number of Monte Carlo simulations were conducted. Since the most frequently used model of the type

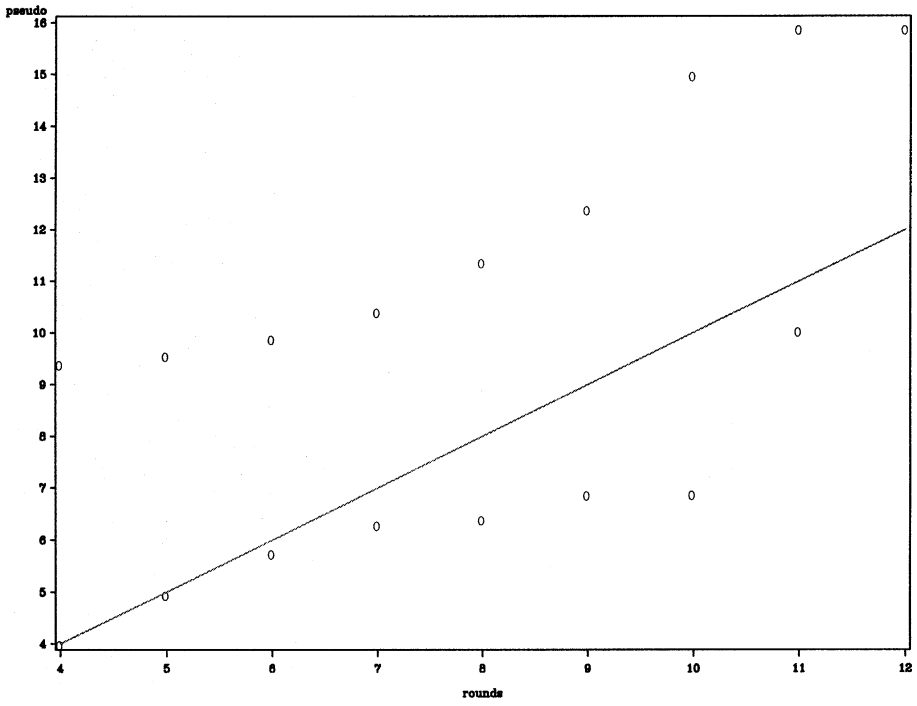

Figure 3. Pseudo-observations for the score of PKA. Upper points correspond to censored observations, lower points correspond to finished games.

(5) is the accelerated failure time model with $h = \log$, attention was focused on this situation. That is, we looked at models for $\mathrm{E}(\log(X))$ using the identity link.

In the simulations $n = 250$ lifetimes were generated from a model given by (5)

$$\log X_i = \beta_0 + \beta_b Z_b + \beta_G Z_G + \varepsilon_i$$

with intercept $\beta_0 = 0$, a binary covariate Z_b and an $N(0,1)$ covariate Z_G. The parameters were $\pi = \mathrm{pr}(Z_b = 1) = 0.5$, and $\beta_b, \beta_G = 0$ or 1. The distribution of the residuals ε_i was either extreme value corresponding to that of X_i being Weibull with shape parameter $\delta = 0.5$, 1 or 2 or log-logistic with scale parameter $\rho = 0.5$, 1 or 2. Letting $\lambda = \exp(\beta_0 + \beta_b Z_b + \beta_G Z_G)$ the survival function for X in the Weibull case is $S(t) = \exp(-\lambda t^{\delta})$ and $\mathrm{E}\varepsilon = -\gamma\delta$, where $\gamma = 0.5772\ldots$ is Euler's constant. For the log-logistic model, $S(t) = (1 + (t/\lambda)^{\rho})^{-1}$ and $\mathrm{E}\varepsilon = 0$. Exponential censoring was superimposed to obtain a censoring percentage of either roughly 25% or 50%. Each combination was repeated 500 times.

Each data set was analyzed in several ways

1. by fitting the correct parametric model (using SAS PROC LIFEREG),
2. by fitting an incorrect parametric model, i.e., a Weibull model for log-logistic data and vice versa (again using SAS PROC LIFEREG),
3. by using the Buckley–James method (using the S-PLUS function bj),
4. by using synthetic data, cf. (6),
5. by using pseudo-observations.

The results for 25% censoring, summarized in Tables 3 and 4, are average $\widehat{\beta}$-values. The bias when estimating $E(\log X)$ using (12) is also given. The empirical standard deviations over the 500 replications (not shown) were similar for all methods and close to the estimated standard errors (for pseudo-observations based on (11)).

It is seen that both the true parametric model and the S-PLUS implementation of the Buckley–James estimator work very well while the approach using the synthetic

Table 3. Average of estimated effects for simulations from the log-logistic distribution with 25% censoring. The three panels correspond to scale parameters $\rho = 2, 1, 0.5$, respectively. The last column is the bias when estimating the mean, μ using (12).

Parameter Combination	Correct Parametric (Log-logistic)	Incorrect Parametric (Weibull)	BJ-Method	Synthetic Data	Pseudo-Observations.	Bias $\widehat{\mu}$
$\beta_b = 0$	0.00	0.00	0.00	0.00	0.00	−0.01
$\beta_G = 0$	0.00	0.00	0.00	0.00	0.00	
$\beta_b = 0$	0.00	0.00	0.00	0.00	0.00	−0.02
$\beta_G = 1$	1.00	0.97	1.00	0.80	0.97	
$\beta_b = 1$	1.00	0.97	1.00	0.83	0.99	−0.01
$\beta_G = 0$	0.00	0.00	0.00	0.00	0.00	
$\beta_b = 1$	1.00	0.97	1.00	0.79	0.98	−0.03
$\beta_G = 1$	1.00	0.97	1.00	0.76	0.96	
$\beta_b = 0$	0.00	0.00	0.00	0.00	0.00	−0.05
$\beta_G = 0$	0.00	0.00	0.01	0.00	0.00	
$\beta_b = 0$	−0.01	−0.02	0.00	0.00	−0.02	−0.07
$\beta_G = 1$	1.01	0.93	1.01	0.80	0.95	
$\beta_b = 1$	1.00	0.92	1.00	0.82	0.96	−0.06
$\beta_G = 0$	0.01	0.01	0.01	0.00	0.01	
$\beta_b = 1$	1.00	0.93	1.00	0.79	0.95	−0.09
$\beta_G = 1$	1.00	0.94	1.00	0.78	0.93	
$\beta_b = 0$	−0.01	0.00	0.00	0.00	−0.01	−0.22
$\beta_G = 0$	0.01	0.00	0.01	0.01	0.01	
$\beta_b = 0$	−0.01	0.00	0.00	0.00	−0.01	−0.23
$\beta_G = 1$	1.00	0.88	1.01	0.80	0.90	
$\beta_b = 1$	1.00	0.88	1.00	0.80	0.90	−0.24
$\beta_G = 0$	0.00	−0.01	0.01	0.01	0.00	
$\beta_b = 1$	1.01	0.89	1.00	0.79	0.90	−0.27
$\beta_G = 1$	1.01	0.91	1.01	0.79	0.91	

Table 4. Average of estimated effects for simulations from the Weibull distribution with 25% censoring. The three panels correspond to shape parameters $\delta = 2$, 1, 0.5, respectively.The last column is the bias when estimating the mean, μ using(12).

Parameter Combination	Correct Parametric (Weibull)	Incorrect Parametric (Log-logistic)	BJ-Method	Synthetic Data	Pseudo-Observations.	Bias$\widehat{\mu}$
$\beta_b = 0$	0.00	0.00	0.00	0.00	0.00	0.00
$\beta_G = 0$	0.00	0.00	0.00	−0.01	0.00	
$\beta_b = 0$	0.00	0.00	0.00	0.00	−0.01	0.00
$\beta_G = 1$	1.00	1.00	1.00	0.84	1.00	
$\beta_b = 1$	1.01	1.00	1.00	0.85	1.00	0.00
$\beta_G = 0$	0.00	−0.01	0.00	−0.01	−0.01	
$\beta_b = 1$	1.00	1.01	1.00	0.84	1.01	0.00
$\beta_G = 1$	1.00	1.00	1.00	0.82	1.00	
$\beta_b = 0$	0.00	0.00	0.00	0.00	0.00	0.00
$\beta_G = 0$	0.00	0.00	0.00	0.00	0.00	
$\beta_b = 0$	0.00	0.00	0.00	0.00	0.00	−0.01
$\beta_G = 1$	1.00	1.01	1.00	0.80	0.99	
$\beta_b = 1$	1.00	1.01	1.00	0.83	1.00	0.00
$\beta_G = 0$	0.00	0.00	0.00	0.00	0.00	
$\beta_b = 1$	1.00	1.01	1.00	0.79	0.98	−0.04
$\beta_G = 1$	0.99	0.99	1.00	0.77	0.95	
$\beta_b = 0$	0.00	0.00	0.00	0.00	0.00	−0.02
$\beta_G = 0$	0.00	0.00	0.00	0.00	0.00	
$\beta_b = 0$	0.00	−0.01	0.00	0.00	−0.01	−0.15
$\beta_G = 1$	1.00	1.02	0.99	0.74	0.89	
$\beta_b = 1$	1.00	1.03	1.00	0.77	0.95	−0.15
$\beta_G = 0$	0.00	0.00	0.00	0.00	0.00	
$\beta_b = 1$	1.00	1.00	1.00	0.73	0.88	−0.29
$\beta_G = 1$	1.00	1.00	0.99	0.70	0.84	

observations (6) only works when the true β is 0. In general, the pseudo-observations work when (12) provides a good estimate of the true $\mathrm{E}(\log X)$. This is for $\rho \neq 0.5$ for the log-logistic distribution and for $\delta \neq 0.5$ for the Weibull, i.e., for the less heavy-tailed distributions. It is seen that pseudo-observations often work slightly better than fitting the incorrect parametric model. For 50% censoring the results (not shown) were similar though (12) more frequently provided inadequate estimates of μ.

3.2. Restricted Mean Survival Time

To study the performance of regression analysis of the restricted mean based on pseudo-observations a small simulation study was set up, as follows: $n = 250$ Weibull distributed life times were generated with scale parameter $\lambda_i = \exp(\beta_b Z_i)$ and shape parameter $\delta = 0.5$, 1 or 2. Here, Z_i is binary with $\mathrm{pr}(Z_i = 1) = 0.5$ and $\beta_b = 0$ or 1. Exponential censoring ($\approx$ 25% or 50%) was superimposed and the

restricted mean life time at τ was estimated for values of τ at the *pth* percentile when $\beta_b = 0$, i.e., $\tau = (-\log(1-p))^{1/\delta}$ for $p = 0.75$ and 0.95. The true value of the restricted mean is then

$$\int_0^\tau \exp(-\lambda t^\delta)dt = \frac{1}{\delta}\lambda^{-1/\delta}\Gamma\left(\frac{1}{\delta}, \lambda\tau^\delta\right),$$

where $\Gamma(a, x)$ is the incomplete gamma function. That is, we have a linear model with intercept $\beta_0 = 1/\delta\Gamma(1/\delta, \tau^\delta)$ and "slope" (effect of Z) $\beta = 1/\delta e^{-\beta_b/\delta}\Gamma(\frac{1}{\delta}, e^{\beta_b}\tau^\delta) - \beta_0$. Each combination was replicated 500 times; results are shown in Table 5. The biases are everywhere quite small (except for one combination). The empirical standard deviations of the estimates were in close agreement with the standard errors based on (11) (not shown).

4. Discussion

Hazard based models have become the primary method of choice for regression analysis of survival data, mainly due to the ease with which right-censoring may be accounted for. However, the simple interpretation of results from classical linear regression models makes analysis of mean survival time appealing but censoring complicates such an analysis. Several methods are available for analysing mean survival and in this paper we have presented an alternative to these techniques based on pseudo-observations. An advantage of our method is that standard programs can be used for the analysis once the pseudo-values are obtained. The method is based on the simple non-parametric estimator for the mean obtained as the integrated Kaplan–Meier estimator and in cases where this estimator is adequate, regression analysis based on pseudo-observations turned out to work well. However, in a

Table 5. Bias when estimating β, the effect of Z (lower panel) on the restricted mean and when estimating the intercept, β_0, (upper panel) at $\tau = P$th percentile in the Weibull distribution with shape parameter δ.

		25% Censoring		50% Censoring	
δ	β_b	$P = 0.75$	$P = 0.95$	$P = 0.75$	$P = 0.95$
0.5	0	−0.003	−0.005	−0.003	−0.005
0.5	1	−0.002	−0.004	−0.003	−0.005
1	0	−0.004	−0.008	−0.005	−0.006
1	1	−0.004	−0.007	−0.004	−0.007
2	0	−0.005	−0.006	−0.006	−0.427
2	1	−0.005	−0.017	−0.006	−0.007
0.5	0	0.004	0.005	0.004	0.053
0.5	1	0.003	0.005	0.004	0.005
1	0	0.006	0.010	0.007	0.012
1	1	0.005	0.010	0.006	0.014
2	0	0.008	0.013	0.009	0.013
2	1	0.007	0.032	0.008	0.117

number of situations, the Buckley–James method provided more stable results than those based on pseudo-observations. An alternative would be to base estimates of the mean on a modified Kaplan–Meier estimator with a parametric tail, e.g., Moeschberger and Klein (1985).

The main advantage of our approach is that it is also directly applicable for performing regression analysis of the restricted mean survival time where few other techniques are available. We studied the use of pseudo-observations in studies of restricted mean survival both in real examples and on simulated data with promising results. One advantage of our approach is that graphical displays may easily be constructed when assessing goodness of fit of the proposed models. Such simple displays have otherwise been lacking in survival analysis. Recent interest in health economics may enhance the applicability of the proposed methods which are also well suited in studies of time to engraftment in bone marrow transplantation.

Acknowledgments

This research was partially supported by Grant no. R01 CA54706-10 from the National Cancer Institute.

References

O. O. Aalen, "A model for non-parametric regression analysis of counting processes," *Springer Lect. Notes Statist.* vol. 2 pp. 1–25, 1980.

O. O. Aalen, "A linear regression model for the analysis of life times," *Statist. Med.* vol. 8 pp. 907–925, 1989.

P. K. Andersen, J. P. Klein, and S. Rosthøj, "Generalized linear models for correlated pseudo-observations, with applications to multi-state models," *Biometrika* vol. 90 pp. 15–27, 2003.

J. D. Buckley and I. R. James, "Linear regression with censored data," *Biometrika* vol. 66 pp. 429–436, 1979.

P. Chen and A. A. Tsiatis, "Causal inference on the difference of the restricted mean lifetime between two groups," *Biometrics* vol. 57 pp. 1030–1038, 2001.

E. Christensen, P. Schlichting, P. K. Andersen, L. Fauerholdt, E. Juhl, H. Poulsen, and N. Tygstrup for The Copenhagen Study Group for Liver Diseases, "A therapeutic index that predicts the individual effects of prednisone in patients with cirrhosis," *Gastroenterology* vol. 88 pp. 156–165, 1985.

Copenhagen Study Group for Liver Diseases, "Sex, ascites and alcoholism in survival of patients with cirrhosis. Effect of prednisone," *New Engl. J. Med.* vol. 291 pp. 271–273, 1974.

D. R. Cox. "Regression models and life-tables" (with discussion). *J. Roy. Statist. Soc. B* vol. 34 pp. 187–220, 1972.

D. M. Dabrowska and K. A. Doksum, "Partial likelihood in transformation models with censored data," *Scand. J. Statist.* vol. 15 pp. 1–24, 1988.

J. P. Fine, Z. Ying, and L.J. Wei, "On a linear transformation model for censored data," *Biometrika* vol. 85 pp. 980–986, 1998.

R. D. Gill. "Large sample behavior of the product-limit estimator on the whole line," *Ann. Statist.* vol. 11 pp. 49–58, 1983.

J. O. Irwin. "The standard error of an estimate of expectation of life, with special reference to expectation of tumourless life in experiments with mice," *J. Hygiene* vol. 47 pp. 188–189, 1949.

E. L. Kaplan and P. Meier, "Non-parametric estimation from incomplete observations," *J. Amer. Statist. Assoc.* vol. 53 pp. 457–481, 562–563, 1958.

T. Karrison, "Restricted mean life with adjustment for covariates," *J. Amer. Statist. Assoc.* vol. 82 pp. 1169–1176, 1987.

T. L. Lai and Z. Ying, "Large sample theory of a modified Buckley–James estimator for regression analysis with censored data,"*Ann. Statist.* vol. 19 pp. 1370–1402, 1991.

S. Leurgans, "Linear models, random censoring and synthetic data," *Biometrika* vol. 74 pp. 301–309, 1987.

J. Li, "An application of lifetime models in estimation of expected length of stay in hospital with complexity and age adjustment," *Stat. Med.* vol. 18 pp. 3337–3344, 1999.

K. -Y. Liang and S. L. Zeger, "Longitudinal data analysis using generalized linear models," *Biometrika* vol. 73 pp. 13–22, 1986.

D. Y. Lin and Z. Ying, "Semi-parametric analysis of the additive risk model," *Biometrika* vol. 81 pp. 61–72, 1994.

M. L. Moeschberger and J. P. Klein, "A comparison of several methods of estimating the survival function when there is extreme right-censoring,"*Biometrics* vol. 41 pp. 253–259, 1985.

Y. Ritov, "Estimation in a linear regression model with censored data," *Ann. Statist.* vol. 18 pp. 303–328, 1990.

P. Schlichting, E. Christensen, P. K. Andersen, L. Fauerholdt, E. Juhl, H. Poulsen, and N. Tygstrup for The Copenhagen Study Group for Liver Diseases, "Identification of prognostic factors in cirrhosis using Cox's regression model," *Hepatology* vol. 3 pp. 889–895, 1983.

A. A. Tsiatis, "Estimating regression parameters using linear rank tests for censored data," *Ann. Statist.* vol. 18 pp. 354–372, 1990.

L. J. Wei, Z. Ying, and D. Y. Lin, "Linear regression analysis of censored survival data based on rank tests," *Biometrika* vol. 77 pp. 845–852, 1990.

Z. Ying, "A large sample study of rank estimation for censored regression data," *Ann. Statist.* vol. 21 pp. 76–99, 1993.

S. L. Zeger and K. -Y. Liang, "Longitudinal data analysis for discrete and continuous outcomes," *Biometrics* vol. 42 pp. 121–130, 1986.

Z. Zheng, "Two methods of estimating the mean survival time from censored samples,"*Sankhyā ser. A* vol. 57 pp. 126–136, 1995.

M. Zhou, "Asymptotic normality of the synthetic data regression estimator for censored survival data," *Ann. Statist.* vol. 20 pp. 1002–1021, 1992.

D. M. Zucker, "Restricted mean life with covariates: Modification and extension of a useful survival analysis method," *J. Amer. Statist. Assoc.* vol. 93 pp. 702–709, 1998.