Pseudo-observations in survival analysis

Per Kragh Andersen and Maja Pohar Perme

Stat Methods Med Res 2010 19: 71 originally published online 4 August 2009
DOI: 10.1177/0962280209105020

Statistical Methods in Medical Research 2010; **19**: 71–99

Pseudo-observations in survival analysis

Per Kragh Andersen Department of Biostatistics, University of Copenhagen, O. Farimagsgade 5, PB 2099, DK 1014 Copenhagen K, Denmark and **Maja Pohar Perme** Department of Biomedical Informatics, University of Ljubljana, Vrazov trg 2, SI-1000 Ljubljana, Slovenia

We review recent work on the application of pseudo-observations in survival and event history analysis. This includes regression models for parameters like the survival function in a single point, the restricted mean survival time and transition or state occupation probabilities in multi-state models, e.g. the competing risks cumulative incidence function. Graphical and numerical methods for assessing goodness-of-fit for hazard regression models and for the Fine–Gray model in competing risks studies based on pseudo-observations are also reviewed. Sensitivity to covariate-dependent censoring is studied. The methods are illustrated using a data set from bone marrow transplantation.

1 Introduction

Survival analysis has developed as an independent area within statistics where methods dealing with incomplete data (right-censoring, left-truncation) are discussed. Examples of such methods include parametric and non-parametric estimation and comparison of survival time distributions and parametric and semi-parametric regression analysis of survival time distributions, often formulated via hazard regression models. Several books on survival analysis, with varying levels of mathematical complexity, are available.[1–3]

Without incomplete data, the survival time X would be observed for all individuals and standard methods for quantitative data could be applied directly for X, or methods for binary outcomes could be applied by dichotomising X as $I(X \leq t)$ for a suitably chosen t. More generally, methods for repeated binary data could be used for a series of indicators, $I(X \leq t_j), j = 1, \ldots, k$. Without incomplete data one could, further, set up regression models for any function, $f(X)$ and check such models using standard graphical methods (scatterplots, residuals) for quantitative (or binary) outcomes.

One way of achieving these goals with censored survival data (and with more general incomplete event history data) is to replace $f(X)$ by the *pseudo-observations* and this article gives a review of such methods and of previous work in the area.[4–11]

The basic idea is simple. If the data were complete, $f(X_i)$ would be observed for each individual i and the expected value $\mathrm{E}(f(X))$ could be estimated by $1/n \sum_i f(X_i)$. Conversely, suppose that the data are incomplete (e.g. some observations are censored and

Address for correspondence: Per Kragh Andersen, Department of Biostatistics, University of Copenhagen, O. Farimagsgade 5, PB 2099, DK 1014 Copenhagen K, Denmark. E-mail: P.K.Andersen@biostat.ku.dk

10.1177/0962280209105020

therefore not all $f(X_i)$ are observed), but a well-behaved estimator, $\widehat{\theta}$, for the expectation

$$\theta = \mathrm{E}(f(X))$$

is available anyway, e.g. the Kaplan–Meier estimator for $S(t) = \mathrm{E}(I(X > t))$. The pseudo-observation for $f(X)$ for individual i, $i = 1, \ldots, n$, is then defined as

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

where $\widehat{\theta}^{-i}$ is the estimator applied to the sample of size $n-1$ obtained by eliminating the i-th individual from the data set. Intuitively the i-th pseudo-observation can be viewed as the contribution of the individual i to the $\mathrm{E}(f(X))$ estimate on the sample of size n. The idea is now to replace the incompletely observed $f(X_i)$ by $\widehat{\theta}_i$, that is

(1) $\widehat{\theta}_i$ may be used as an outcome variable in a generalised linear regression model with some link function g:

$$g(\mathrm{E}(f(X) \mid Z)) = \beta_0 + \sum \beta_j Z_j,$$

or

(2) $\widehat{\theta}_i$ may be used to compute residuals or in a scatterplot when assessing model assumptions.

Regardless of the application, the pseudo-observations $\widehat{\theta}_i$ will always be used for all n subjects and not only for those where $f(X_i)$ was unobserved.

The structure of the article is as follows. In Section 2, we explain the rationale behind pseudo-observations, review some of their properties and present examples to be used in later sections. In Section 3, we demonstrate how pseudo-observations may be used for analysing regression models for parameters of the form $\mathrm{E}(f(X))$, cf. 1. above. In Section 4, we discuss how pseudo-observations may be used for assessing goodness-of-fit, both for models introduced in Section 3, but also for standard models for survival data, including the Cox proportional hazards model, cf. 2. above. Section 5 contains some discussion and concluding remarks.

2 Definition of pseudo-observations

Let X be a random variable, which may be a scalar, a vector $X = (X_1, \ldots, X_k)$, or a process $X(t)$, and let θ be a parameter of the form $\theta = \mathrm{E}(f(X))$. Here, f and thereby θ may be multivariate. Our main interest will be in survival or event history analysis where pseudo-observations present a general approach that can be used analogously in several situations. To illustrate this, we shall present the following examples.

Example 1. Survival probabilities
The starting point is here a survival time X and the survival probability, $S(t_0) = \mathrm{E}(I(X > t_0))$ at a fixed time point, t_0, at a number of fixed time points, $t_1, \ldots, t_k$, i.e. $S(t_j) = \mathrm{E}(I(X > t_j))$, or at all time points $t \geq 0$. Thus, in this example the function is $f(X) = f_t(X) = I(X > t)$ and the parameter θ is $S(t)$.

Example 2. Restricted mean survival time
The starting point is again a survival time, X, and the τ-restricted mean

$$\mu_\tau = \mathrm{E}(X \wedge \tau)$$

for some fixed $\tau > 0$. That is, $f(X) = f_\tau(X) = X \wedge \tau$ and θ is μ_τ. Note that, since

$$X \wedge \tau = \int_0^{X \wedge \tau} 1 \mathrm{d}t = \int_0^\tau I(X > t) \mathrm{d}t,$$

we have $\mathrm{E}(X \wedge \tau) = \int_0^\tau S(t) \mathrm{d}t$. Discussions on when and how to use the restricted mean have been presented previously.[5]

Example 3. Competing risks cumulative incidences
The starting point is here the competing risks model with, say, two causes of failure, that is the multi-state process $X(t)$ taking values 0 ('alive at t'), 1 ('failure from cause 1 before t') or 2 ('failure from cause 2 before t'), cf. Figure 1. Here, interest may focus on the cumulative incidence of failure from cause 1 at a fixed time point, t_0, that is $C_1(t_0) = \mathrm{E}(I(X(t_0) = 1))$. As in the first example, several time points, $t_1, \ldots, t_k$ or all time points $t > 0$ may be considered. Representing the process by the failure time X and the cause-of-death indicator, $D = 1, 2$, the cumulative cause 1 incidence may be written in the form

$$C_1(t) = \mathrm{E}(I(X \leq t, D = 1)) = \int_0^t S(u-)\alpha_1(u) \mathrm{d}u,$$

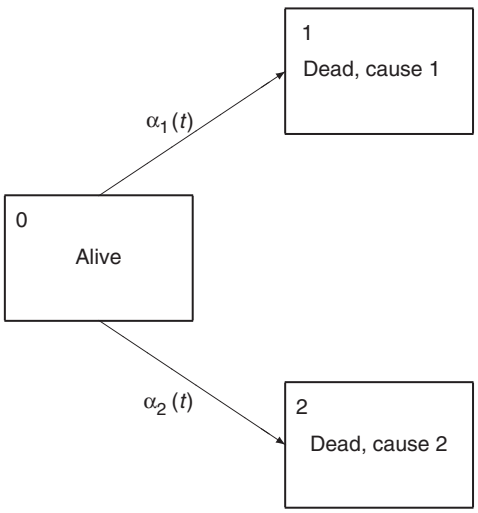

Figure 1 The competing risks model with two causes of failure.

where $S(u) = \mathrm{P}(X > u)$ is the overall survival probability (that is, $\mathrm{P}(X(u) = 0)$) and $\alpha_1(u)$ is the cause specific hazard for failures from cause 1. Thus, here $f(X(\cdot)) = f_t(X(\cdot)) = I(X(t) = 1)$ and θ is $C_1(t)$.

Examples 1–3 are those that we will work with in detail in this article. A final example to be mentioned occasionally is the following.

Example 4. The illness-death model

Consider the three state illness-death model[12] depicted in Figure 2. That is, we study the multi-state process, $X(t)$ taking values 0 ('disease-free at t'), 1 ('alive with disease at t') or 2 ('death before t'). In this model interest may focus on the probability of being diseased at time t, $\mathrm{E}(I(X(t) = 1))$, i.e. $f(X(\cdot)) = I(X(t) = 1)$ and θ is the state occupation probability, $Q_1(t) = \mathrm{P}(X(t) = 1)$.

Suppose that i.i.d. copies $X_1, \ldots, X_n$ of X are observed, possibly subject to independent right-censoring (precise specification of the available data to be given in the examples below). Suppose further that, based on these n observations, an (approximately) unbiased estimator, $\widehat{\theta}$ of $\theta = \mathrm{E}(f(X))$ is available. We then define the pseudo-observation for $f(X_i)$ as

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

with $\widehat{\theta}^{-i}$ being the estimator applied to the sample of size $n-1$ obtained by eliminating subject i from the total sample.

The idea is the following: for complete data, i.e. data with no censoring, $\theta = \mathrm{E}(f(X))$ could be estimated unbiasedly by the average $(1/n)\sum_i f(X_i)$ in which case the i-th pseudo-observation would simply be $f(X_i)$. With censoring we then replace the, possibly unobserved, $f(X_i)$ by the pseudo-observation $\widehat{\theta}_i$. That is, when analysing regression models for θ, i.e. models for $\mathrm{E}(f(X) \mid Z)$, the pseudo-observation $\widehat{\theta}_i$ is used as response variable. Similarly, when assessing goodness-of-fit of such a model using scatterplots or residual plots, the pseudo-observation $\widehat{\theta}_i$ is again used as response variable. Note that we will always use the pseudo-observation as response variable, also for those subjects where $f(X_i)$ was actually observed.

Following the definition, the pseudo-observations for each of the examples can be calculated as follows.

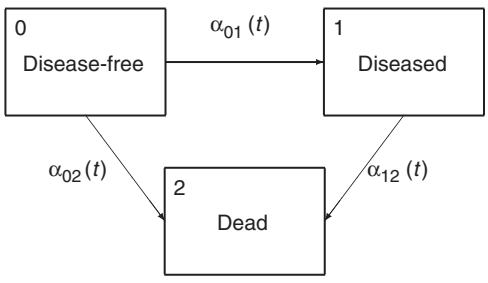

Figure 2 The illness-death model.

Example 1. Survival probabilities (ctd)
In this example the data are the usual right-censored survival data, i.e. $(\widetilde{X}_i, D_i)$, $i = 1, \ldots, n$, where $\widetilde{X}_i$ equals the true survival time X_i when $D_i = 1$ and it equals the right-censoring time when $D_i = 0$. These data can be summarised as the counting process

$$N(t) = \sum_i I(\widetilde{X}_i \leq t, D_i = 1),$$

giving the number of observed failures in $[0, t]$, and

$$Y(t) = \sum_i I(\widetilde{X}_i \geq t),$$

the number of subjects observed to be still at risk at time $t-$. The estimator for $\theta = S(t)$ is the Kaplan–Meier estimator,

$$\widehat{S}(t) = \prod_{u \leq t} \left(1 - \frac{\mathrm{d}N(u)}{Y(u)}\right),$$

which is approximately unbiased under independent censoring.[2] The i-th pseudo-observation is then

$$\widehat{S}_i(t) = n \cdot \widehat{S}(t) - (n-1)\widehat{S}^{-i}(t). \tag{1}$$

Example 2. Restricted mean life time (ctd)
The data are the same as in Example 1 (above) and the integral of the Kaplan–Meier estimator

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

estimates the τ-restricted mean $\theta = \mathrm{E}(X \wedge \tau)$ approximately unbiasedly. The i-th pseudo-observation will then be

$$\widehat{\mu}_{\tau i} = n \int_0^\tau \widehat{S}(t)\mathrm{d}t - (n-1) \int_0^\tau \widehat{S}^{-i}(t)\mathrm{d}t = \int_0^\tau \widehat{S}_i(t)\mathrm{d}t \tag{2}$$

with $\widehat{S}_i(t)$ given by (1).

Example 3. Competing risks cumulative incidences (ctd)
The data in this example are $(\widetilde{X}_i, D_i), i = 1, \ldots, n$, with $\widetilde{X}_i$ being the, possibly right-censored, failure time and where $D_i = 0$ still indicates censoring while $D_i = j, j = 1, 2,$

if subject i was observed to die from cause j. The counting process representation of these data is

$$N_j(t) = \sum_i I(\tilde{X}_i \leq t, D_i = j), \quad j = 1, 2,$$

counting cause j failures in $[0, t]$, and $Y(t)$ which, as above, gives the number of subjects still at risk at time $t-$. The cumulative incidence for cause 1 is estimated by the Aalen–Johansen estimator[2]

$$\widehat{C}_1(t) = \int_0^t \widehat{S}(u-)\mathrm{d}\widehat{A}_1(u),$$

where $\widehat{A}_1(t) = \int_0^t \mathrm{d}N_1(u)/Y(u)$ is the Nelson–Aalen estimator[2] for the cumulative cause-specific hazard $A_1(t) = \int_0^t \alpha_1(u)\mathrm{d}u$ for cause 1 failures. The i-th pseudo-observation will be denoted $\widehat{C}_{1i}(t)$.

2.1 Properties of pseudo-observations

We will first illustrate how $\widehat{\theta}_i$ may look like for each of the three examples presented. Figure 3 shows the case of the survival indicator $I(X_i > t)$. In the case of no censoring in the data set, formula (1) simplifies into pseudo-observations being equal to the survival experience of the individual i: $S_i(t) = I(X_i > t)$ is equal to 1 while the subject is still alive and drops down to 0 when he dies (Figure 3(a)). Figures 3(b) and (c) present examples for two individuals from a sample with 25% censoring ($n = 1000$). Pseudo-observations are still defined at all times for all (censored and uncensored) individuals. As pseudo-observations for each individual depend on the whole sample and are calculated using the Kaplan–Meier estimate (1), their value changes at every event time and is constant in between. We can observe that pseudo-observations for individuals still at risk at a certain time point all have the same value (above 1). This is due to the fact that omitting them in $\widehat{S}^{-i}$ lowers the risk set, regardless of the censoring status and the actual event time.

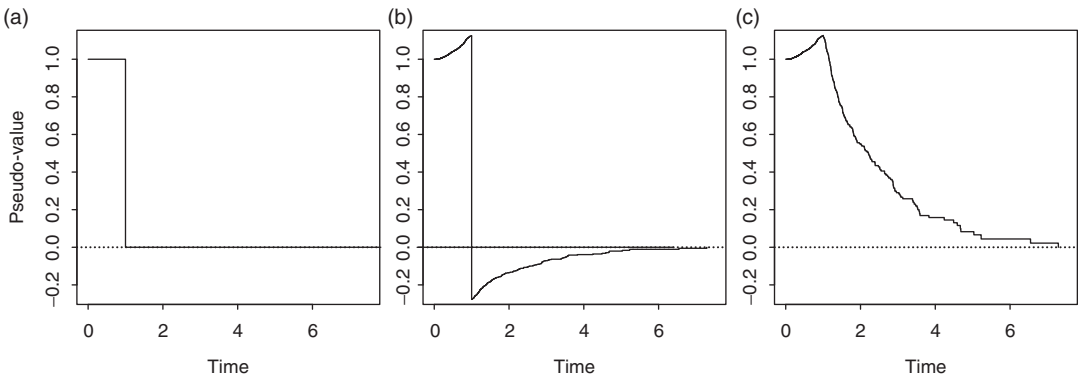

Figure 3 The pseudo-observations for the survival function in time. (a) The pseudo-observations for an individual with $X_i = 1$ in a data set with no censoring; (b) the pseudo-observations for an individual from a censored data set, who experienced an event at time $X_i = 1$; (c) the pseudo-observations for an individual, censored just after $X_i = 1$.

As the risk set is diminishing in time, the discrepancy between $\widehat{S}$ and $\widehat{S}^{-i}$ is increasing causing the pseudo-observations to increase as well. The end of follow-up causes a jump in case of an event (Figure 3(b)) and a turning point in case of censoring (Figure 3(c)). Note that the turning point in the latter figure is actually at the last event time before this individual was censored, since the pseudo-observations (like the Kaplan–Meier estimates) only change at event times.

In Figures 4 and 5 we display pseudo-observations for the τ-restricted mean life time $X_i \wedge \tau$. In this case, one pseudo-observation is calculated for each individual. In the case of no censoring, the pseudo-observation is equal to the event time, X_i, of individual i if this is $\leq \tau$, see Figure 4, otherwise it equals τ. Figure 5 presents situations with various degrees of censoring and choices of τ. The pseudo-observations for the survival function are above 1 until the time of event or censoring X_i (Figure 3(b) and (c)) and the integral (2) up to this time is larger than X_i:

$$\int_0^{X_i} \widehat{S}_i(t)\mathrm{d}t > \int_0^{X_i} 1\mathrm{d}t = X_i.$$

Since the pseudo-observations for the survival function of a censored individual remain positive also after the time of censoring X_i (Figure 3(c)), the pseudo-observations for the restricted mean for such an individual (grey dots) are always larger than X_i and increase with τ. On the other hand, the pseudo-observations for events (black dots) are mostly smaller than the follow-up times and decrease with τ.

Finally, Figures 6 and 7 present pseudo-observations for the competing risks cause 1 cumulative incidence function $I(X_i \leq t, D_i = 1)$. When no censoring is present, the pseudo-observations are equal to the failure indicator of individual i: $C_{1i}(t) = I(X_i(t) = 1)$, i.e. equal to 0 until an event happens and then either remaining at zero or moving to 1, depending on the failure cause.

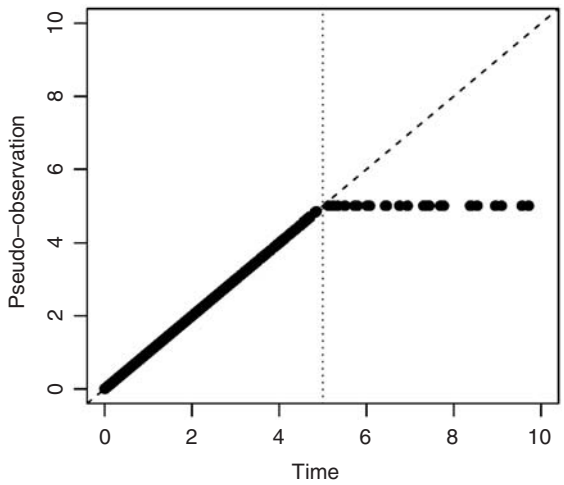

Figure 4 The pseudo-observations for the restricted mean with respect to the follow-up time of each individual for the case of no censoring. The truncation time is 5 years (dotted line), the dashed line marks the line of equality.

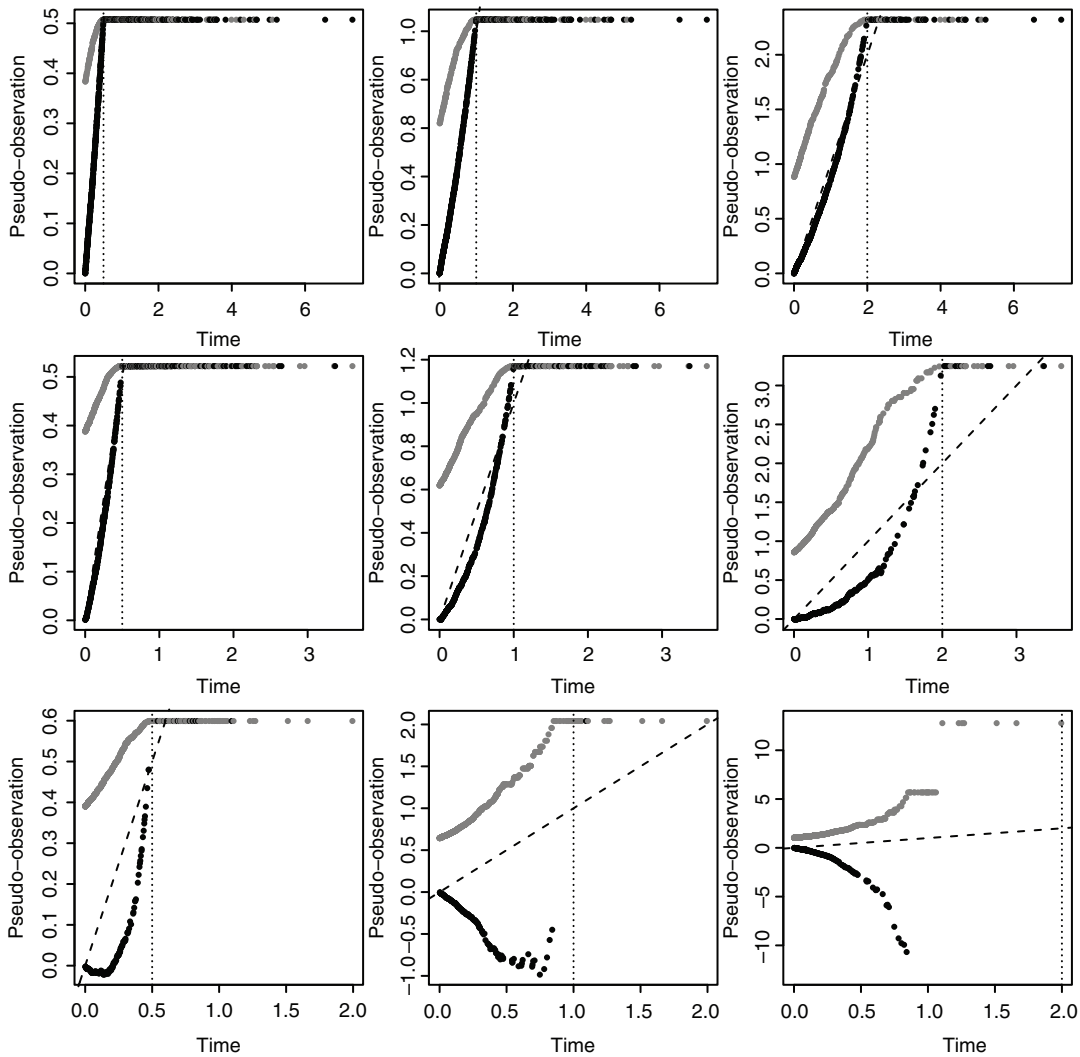

Figure 5 The pseudo-observations for the restricted mean with respect to the follow-up time of each individual for various degrees of censoring (top = 25%, middle = 50%, bottom = 75%) and various choices of τ (left = 0.5, middle = 1, right = 2). The grey dots denote censored and the black ones uncensored individuals; truncation time is denoted by a vertical dotted line and the line of equality is marked by the dashed line. One point (1.1, −44) is omitted on the bottom right graph as its value is too far out of range.

In the case of censoring, pseudo-observations tend to be negative at first and then jump above 1 in case of failing from the cause in question or remain negative (and decrease) in the case of failing from the other cause, see Figures 7(a) and 7(b). If an individual is censored, the pseudo-observations start increasing at the first next failure corresponding to the cause in question, see Figure 7(c).

Unbiasedness of $\widehat{\theta}$ and the fact the observations are i.i.d. immediately ensure that the (marginal) expectation of the i-th pseudo-observation equals θ:

$$\mathrm{E}(\widehat{\theta_i}) = n \cdot \theta - (n-1)\theta = \theta.$$

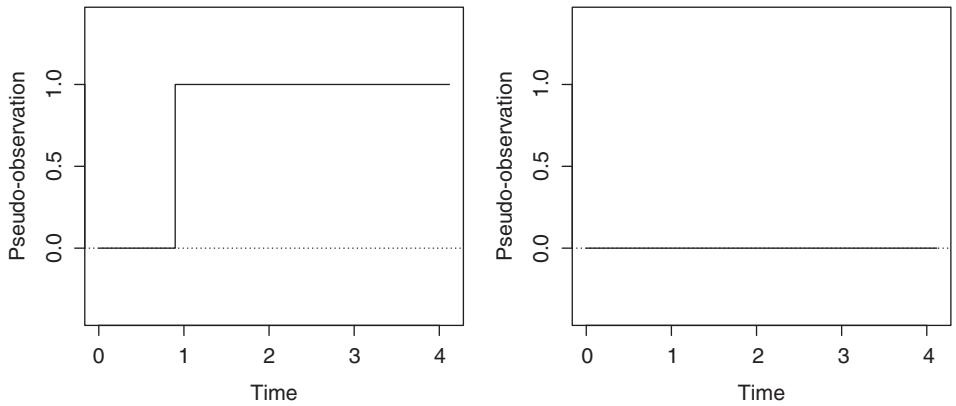

Figure 6 The pseudo-observations for the cumulative incidence function (for cause 1) in time for the case of no censoring. (a) Individual dying from cause 1; (b) Individual dying from cause 2.

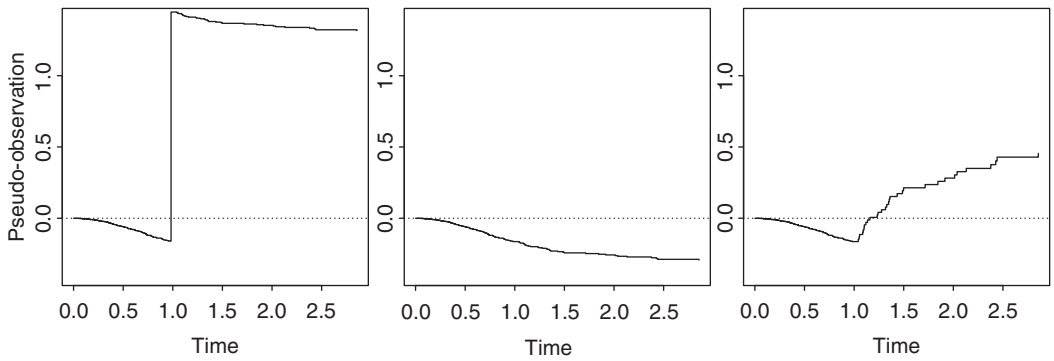

Figure 7 The pseudo-observations for the cumulative incidence function (for cause 1) in time for a censored data set. (a) Individual failing from cause 1; (b) Individual failing from cause 2; (c) Censored individual.

Tukey[13] conjectured that $\widehat{\theta}_i, i = 1, \ldots, n$ are approximately i.i.d. A version of this was, for the special case of the competing risks cumulative incidence function (Example 3), proved by Graw *et al.*[11] using results from influence functions. These authors showed that, for any fixed number m, the first m pseudo-observations $\widehat{C}_{1i}(t), i = 1, \ldots, m$, (or any other subset of size m) are approximately i.i.d. when the sample size, n, tends to infinity. By this result, Tukey's conjecture also holds for the Kaplan–Meier estimator (Example 1) since, with no competing risks, the estimated cumulative cause 1 incidence function function $\widehat{C}_1(t)$ reduces to $1 - \widehat{S}(t)$. This, finally, implies the conjecture for the restricted mean of Example 2, see Equation (2).

Similarly, Graw *et al.*[11] studied a regression situation where i.i.d. covariate vectors $Z_i, i = 1, \ldots, n$ are given. For the cumulative incidence function they showed that the conditional expectation of the i-th pseudo-observation given covariates, $\mathrm{E}(\widehat{C}_{1i}(t) \mid Z_i)$, in large samples, i.e. as $n \to \infty$, equals its theoretical counterpart $\mathrm{E}(I(X_i \leq t, D_i = 1) \mid Z_i)$. By arguments as above, this property then transfers to the Kaplan–Meier estimator and

to the restricted mean. We will need such a property in the next section when regression models based on pseudo-observations are to be discussed.

Results like those obtained by Graw *et al.*[11] are not yet available for pseudo-observations computed in a general situation.

2.2 Dependent censoring

All of these nice properties rely on the fact that the Kaplan–Meier estimator is (approximately) unbiased for the marginal survival function $S(t) = \mathrm{P}(X > t)$ (and on similar properties for $\widehat{\mathrm{C}}_1(t)$). That is the case under independent censoring, in particular when the censoring does not depend on covariates, see e.g. Andersen *et al.*,[2] Sections IV.3-4, who showed that the bias decreases exponentially with the sample size, n. However, if censoring does depend on covariates then these properties may be lost. We studied this problem in a series of Monte Carlo simulations.

Survival data ($n = 500$) were generated in two groups from the Cox model $\alpha(t \mid Z) = \alpha_0(t)\exp(\beta Z)$ with $\alpha_0(t) \equiv 1$, $\beta = 1$ and Z binary 0,1 with either $\mathrm{P}(Z = 1) = 0.5$ or 0.9. Censoring was either independent, exponential (~50%) or dependent following the same Cox model as the survival times. Results based on 1000 replications of the situation with $\mathrm{P}(Z = 1) = 0.5$ are shown in Table 1. The upper panel shows that, under independent censoring, the average of the pseudo-observations for the survival indicator computed at times 0.5, 1 or 1.5 is close to the true value of $S(t)$. In fact, Stute and Wang[14] showed that for values of t less than the last failure time, the average of the pseudo-observations equals the Kaplan–Meier estimator. Also, in the upper panel, it is seen that for the situation with dependent censoring, the pseudo-observations are biased. This is because, as mentioned above, the Kaplan–Meier estimator is no longer unbiased for the marginal survival function when censoring is dependent.

If censoring depends on a categorical covariate, Z, with values, say, $1, 2, \ldots, m$, then the Kaplan–Meier estimator based on all data may be replaced by the mixture:

$$\widehat{S}_M(t) = \sum_{j=1}^{m} p_j \widehat{S}_j(t), \tag{3}$$

where p_j is the fraction of subjects with $Z = j$ and $\widehat{S}_j(t)$ is the Kaplan–Meier estimator based on group j, only. For this estimator the i-th pseudo-observation, when subject i

Table 1 Simulation results. Upper panel: Kaplan–Meier estimates. Lower panel: 'mixture estimator' (3)

	$t = 0.5$		$t = 1$		$t = 1.5$	
	$S(t)$	*SD*	$S(t)$	*SD*	$S(t)$	*SD*
True value	0.432		0.217		0.120	
Independent censoring	0.431	0.029	0.218	0.031	0.122	0.035
Dependent censoring	0.465	0.029	0.263	0.031	0.158	0.031
Independent censoring	0.431	0.029	0.217	0.030	0.121	0.035
Dependent censoring	0.431	0.030	0.214	0.036	0.122	0.039

belongs to group j, is simply that based on $\widehat{S}_j(t)$. In this situation, the value of t may well exceed the range of observations in a given group, j. In that case, the last value of $\widehat{S}_j(t)$ may be carried forward to define its value in t.

In the lower panel of Table 1, pseudo-observations were based on the mixture estimator (3). It is seen that the bias then disappears while the standard deviation tends to increase slightly. Similar results (not shown) were found for the situation where $P(Z = 1) = 0.9$ and for the case where the censoring times are uniformly distributed and dependent on covariates.

3 Regression models based on pseudo-observations

When regressing on covariates in survival analysis one may be interested in several different outcomes. This includes modelling the survival experience with, e.g. a Cox model, but also examples like the restricted mean where no standard methods for modelling in the presence of censored data exist. Pseudo-observations provide a common approach to various kinds of models by replacing the incompletely observed outcome and then fitting using generalised estimating equations. To illustrate the method in its generality, we shall consider all the examples mentioned in Section 2, though the real practical value lies in the cases where no other standard methods apply.

3.1 Specification of the models

As exemplified in the previous section, let X be a scalar or a vector $X = (X_1, \ldots, X_k)$ random variable, or consider a stochastic process $X(t)$. Let θ be a parameter of the form $\theta = E(f(X))$. Here, f and thereby θ may be multivariate (and even infinite dimensional, see below). Finally, let $Z = (Z_1, \ldots, Z_p)$ be a vector of (time-fixed) covariates on which we want to regress $f(X)$, that is, we will study a model for $E(f(X) \mid Z)$. We will restrict attention to the class of generalised linear models where, for a given link function, g, we have:

$$g(E(f(X) \mid Z)) = \beta_0 + \sum \beta_j Z_j.$$

Examples of this situation are frequent and include ordinary linear regression (where X is a quantitative outcome variable and the link is the identity, $g(x) = x$) and logistic regression (where X is binary and $g(x) = \log \frac{x}{1-x}$). In both of these examples, f is the identity (that is, $E(X)$ is modelled directly) and thereby, f is univariate. Examples from survival analysis with a univariate f include the Cox proportional hazards regression model in a single time point, t_0 (Example 1 above) where $f(X) = I(X > t_0)$ and, thereby, $\theta = S(t_0)$. The link function is here the cloglog-function $g(x) = \text{cloglog}(x) = \log(-\log(x))$. The generalised linear model then becomes

$$\log(-\log(S(t_0))) = \log A_0(t_0) + \sum_j \beta_j Z_j, \tag{4}$$

where $A_0(t) = \int_0^t \alpha_0(u)\mathrm{d}u$ is the cumulative baseline hazard in the Cox model $\alpha(t \mid Z) = \alpha_0(t)\exp(\sum_j \beta_j Z_j)$. This model can be extended to a joint proportional hazards model for several time points $t_1, \ldots, t_k$:

$$\log(-\log(S(t_\ell))) = \log A_0(t_\ell) + \sum_j \beta_j Z_j, \quad \ell = 1, \ldots, k,$$

where $f(X) = (f_{t_1}(X), \ldots, f_{t_k}(X))$ (with $f_{t_j}(X) = I(X > t_j)$) is k-dimensional. Finally, when considering a model for *all* time points:

$$\log(-\log(S(t))) = \log A_0(t) + \sum_j \beta_j Z_j, t > 0,$$

f is even infinite-dimensional. For the τ-restricted mean of Example 2, a linear model

$$\mathrm{E}(X \wedge \tau \mid Z) = \beta_0 + \sum_j \beta_j Z_j$$

or a log-linear model

$$\log(\mathrm{E}(X \wedge \tau \mid Z)) = \beta_0 + \sum_j \beta_j Z_j$$

could be of interest while, in the competing risks Example 3, the Fine–Gray model[15] could be studied. Here

$$\log(-\log(1 - C_1(t))) = \beta_0(t) + \sum_j \beta_j Z_j, \quad t > 0.$$

Also, for the illness-death model of Example 4, a logistic or a cloglog model for $Q_1(t)$ could be of interest.

3.2 Model fitting using pseudo-observations

To fit such models based on censored data, pseudo-observations may be used. For individual i, the pseudo-observation $\widehat{\theta}_i$ is k-dimensional (k time points), $\widehat{\theta}_i = (\widehat{\theta}_{ij}, j = 1, \ldots, k)$ where k may be 'large'. Thus, in the Cox model or the Fine–Gray model for all time points one would typically have a value for all observed event times, that is, the times at which the pseudo-observations change, cf. Figures 3 and 7. In these cases $\widehat{\theta}_{ij}$ represents $f_{t_j}(X_i)$ and the model is $g(\mathrm{E}(f_{t_j}(X_i) \mid Z_i) = \beta^\top Z^*_{ij}$. Here, the vector Z^*_{ij} will include indicators of the time points as well as the covariates Z_i, i.e. $Z^*_{ij} = (I(t_\ell = t_j), \ell = 1, \ldots, k; Z_i)$.

Estimates of the ($k+p$-dimensional) βs may be based on the estimating equations

$$\sum_i \left(\frac{\partial}{\partial \beta} g^{-1}(\beta^\top Z_i^*) \right)^\top V_i^{-1} (\widehat{\theta}_i - g^{-1}(\beta^\top Z_i^*)) = \sum_i U_i(\beta) = U(\beta) = 0. \qquad (5)$$

Here, V_i is a working covariance matrix and $g^{-1}(\beta^\top Z_i^*)$ is short for the k-vector with elements $g^{-1}(\beta^\top Z_{ij}^*), j = 1, \ldots, k$. A sandwich estimator can be used to estimate the variance of $\widehat{\beta}$. Let

$$I(\beta) = \sum_i \left(\frac{\partial g^{-1}(\beta^\top Z_i^*)}{\partial \beta} \right)^\top V_i^{-1} \left(\frac{\partial g^{-1}(\beta^\top Z_i^*)}{\partial \beta} \right),$$

$$\widehat{\mathrm{var}}(U(\beta)) = \sum_i U_i(\widehat{\beta})^\top U_i(\widehat{\beta}),$$

then

$$\widehat{\mathrm{var}}(\widehat{\beta}) = I(\widehat{\beta})^{-1} \widehat{\mathrm{var}}(U(\beta)) (I(\widehat{\beta})^{-1}).$$

Alternative estimators of the variance of $\widehat{\beta}$ can be found by using a non-parametric bootstrap technique, i.e. by resampling with replacement from the original records.

Using GEE results of Liang and Zeger,[16] conditions for the estimators of β to be asymptotically normal were presented by Graw *et al.*[11] for the case of pseudo-observations from the cumulative incidence function (Example 3). By arguments similar to those presented in the previous section these asymptotic results also hold for the survival function (Example 1) and for the τ-restricted mean (Example 2). The results of Graw *et al.*[11] cover the case where the number of time points, k is finite. Results for all values of $t > 0$ as well as for general situations, including Example 4, are not yet available.

Once the pseudo-observations have been computed, estimators of β can be obtained by using standard software for generalised estimating equations (GEE) such as `PROC GENMOD` in `SAS` or the `geese` function in `R` (see Klein *et al.*[9] for a discussion of this point and for a presentation of `SAS MACRO`s and `R`-functions for computation of the pseudo-observations in Examples 1–3).

Notice that, when there is no censoring, the estimators of θ in all our examples reduce to simple relative frequencies and in this case the pseudo-observations for subject i are simply $f_{t_j}(X_i), j = 1, \ldots, k$. For the uncensored situation, standard methods for inference are available, including linear regression of $T \wedge \tau$ and a simple version of the Fine–Gray model[15] for the cumulative incidence function.

As for other applications of GEE, there is a choice of working covariance matrix, V_i, to be made. Choosing V_i to resemble the true covariance as closely as possible will increase the efficiency of the estimators.[16] This was to some extent confirmed by the simulations reported by Andersen and Klein[7] for the competing risks model, Example 3. Also, this simulation study investigated the impact of the choice of the number of time points, k, which also has to be made by the user. The results suggested that increasing k did not improve efficiency considerably and as few as five time points (equally spread on

the event time scale) in most cases worked quite well. No study of the performance of the method when using all time points is available. Note, however, the close resemblance between this situation and that considered in the temporal process regression approach of Fine *et al.*[17]

Further simulation studies were reported in Andersen *et al.* (2003)[4] for the three-state illness-death model of Example 4, in Andersen *et al.* (2004)[5] for the restricted mean survival time of Example 2, and in Klein *et al.*[8] for the survival function in a single time point, Example 1.

3.3 Dependence on censoring

Independent censoring is a requirement for the GEE to be unbiased. We illustrate this by continuing the simulation study from the previous section. For the situations studied there, pseudo-observations for the survival function were computed at time points 0.25, 0.5, 1 and 1.5 and used as outcome variables in the GEE for the model (4). Table 2 shows the results. While estimates are always unbiased under independent censoring, dependent censoring creates a bias when using pseudo-observations based on the biased Kaplan–Meier estimator. This bias, however, disappears when using the mixture estimator (3) but the SD seems to increase under dependent censoring.

Table 2 also shows that while estimates for the Cox model may be based on pseudo-observations these estimates will be more dispersed than those based on the standard partial likelihood method. So, obviously, we do not recommend to use pseudo-observations in cases where standard methods apply. The strength of the method is to provide a way of inference in cases where standard methods may not be available. These include analysis of the survival function in a single time point, the restricted mean survival time and (to some extent) the competing risks cumulative incidence function. A further strength, to be expanded upon in Section 4, is that pseudo-observations may be used for graphical goodness-of-fit assessment, also in standard models like the Cox model[10] or the Fine–Gray model.

Some model assumptions, including linearity, constant effects over time and absence of certain interactions may be tested in the usual way by adding appropriate terms to the models. Thus, constant effects over time may be tested by simply adding an interaction between a given covariate and time, while linearity may be tested by adding, e.g. restricted cubic splines to the linear model terms. In Section 4, we will also illustrate this use of pseudo-observations for numerical model checking.

Table 2 Estimates of the regression coefficient based on GEE for pseudo-observations

	Independent		Dependent	
	β	*SD*	β	*SD*
True value	1		1	
Cox model	1.008	0.133	1.010	0.145
Pseudo-obs., Kaplan–Meier	1.019	0.143	0.859	0.137
Pseudo-obs., 'mixture' estimator (3)	1.012	0.143	0.998	0.161

3.4 Illustration

We will illustrate the techniques using a bone marrow transplant data set from The Center for International Blood and Marrow Transplant Research (CIBMTR). The CIBMTR is comprised of clinical and basic scientists who confidentially share data on their blood and bone marrow transplant patients with CIBMTR Data Collection Center located at the Medical College of Wisconsin. The CIBMTR is a repository of information about results of transplants at more than 450 transplant centres worldwide. The example data set consists of patients who received HLA-identical sibling transplant from 1995 to 2004 for acute myelogenous leukemia (AML) or acute lymphoblastic leukemia (ALL) and were transplanted in first complete remission. All patients received bone marrow transplantation or peripheral blood stem cell transplantation. The infants aged less than 2 years old and all patients who received umbilical cord blood transplants were excluded as risk factors are likely to vary in this group. Table 3 presents a description of the illustrative data set.

3.4.1 Survival probabilities

We consider the 5-year overall survival probability and study how covariates are associated with this outcome using pseudo-observations based on $\widehat{S}(5)$ via a cloglog or a logit link function. Notice once again that, without censoring, the survival indicator $I(X_i > 5)$ would have been observed for all subjects, i, and a standard binary regression model could have been applied. Table 4 shows the results. There are no major differences in the conclusions based on the two different link functions. AML patients have lower 5-year survival probabilities ($\beta > 0$) than ALL patients, the survival probability decreases with increasing age and increases with increasing Karnofsky score, while patients treated only with bone marrow have lower survival probabilities than those treated with peripheral blood stem cells.

3.4.2 Restricted mean survival

For the 5-year restricted mean survival time, pseudo-observations were based on $\int_0^5 \widehat{S}(t)\mathrm{d}t$. The results from a regression model with the identity link, presented in Table 5, are qualitatively compatible with those based on the 5-year survival probability though, formally, the effect of graft type is not significant. Note the simple interpretation of the

Table 3 Description of data from 2009 patients who underwent bone marrow transplantation

Age (mean, SD)	31.9	15.4
Female sex	896	44.6%
Disease type AML	1406	70.0%
Graft type BM only	1153	57.4%
Karnofsky score (mean, SD)	91.3	9.1 (25 missing values)
Relapse	259	12.9%
Death	737	36.7%
Death and relapse	232	89.6% of patients with relapse

Table 4 Regression model for the 5-year survival probability

	Cloglog link			Logit link		
Covariate	β	SE	p	β	SE	p
Disease (AML vs. ALL)	0.384	0.084	<0.001	0.481	0.110	<0.001
Age (per 10 years)	0.149	0.029	<0.001	0.185	0.037	<0.001
Graft type (BM only vs. other)	0.207	0.083	0.013	0.255	0.107	0.017
Karnofsky score (per 10 points)	−0.127	0.041	0.002	−0.158	0.055	0.004

Table 5 Regression model the 5-year restricted mean

Covariate	β	SE	p
Disease (AML vs. ALL)	−0.427	0.103	<0.001
Age (per 10 years)	−0.168	0.033	<0.001
Graft type (BM only vs. other)	−0.149	0.101	0.138
Karnofsky score (per 10 points)	0.160	0.052	0.002

effects, e.g. AML patients (with given values of age, graft type, and Karnofsky) live on average 0.427 years shorter than similar ALL patients during the first five years after transplantation.

Model checking based on the pseudo-observations for both examples will be illustrated in the next section.

3.4.3 Competing risks cumulative incidence function

Pseudo-observations for the cumulative incidence functions for both relapse and death in remission were computed at four time points 0.23, 0.5, 1.21 and 5 (same time points for both outcomes, the first three chosen as quartiles in the distribution of all observed event times). Models were fitted to these observations including the same covariates as in the previous two examples and using both the cloglog and the logit link functions. Table 6 shows the results. No major differences are seen between the two link functions. For both outcomes, AML patients have higher risks than those with ALL and age mainly affects the death risk while, in fact, older patients tend to have slightly lower risks of relapse. Graft type has no association with either outcome and increasing Karnofsky score is associated with lower death risks but does not seem to be related to the risk of relapse. Also, for this analysis, we will return to an examination of goodness-of-fit in the next section. For comparison we also analysed the data using the Fine–Gray model where no choice of time points is needed. The results, presented in Table 7, are roughly similar to those in Table 6 using the cloglog link though standard errors were slightly smaller. We also fitted the models to the pseudo-observations computed at *all* event times, Table 7, however, to avoid too many parameters the baseline was modelled using a five degree-of-freedom spline function. Again, the results are close to those based on only four points.

Table 6 Regression models for the competing risks cumulative incidences of relapse and death in remission

	Cloglog			Logit		
Covariate	β	SE	p	β	SE	p
Relapse:						
Disease (AML vs. ALL)	0.445	0.134	<0.001	0.475	0.144	<0.001
Age (per 10 years)	−0.096	0.047	0.039	−0.105	0.05	0.036
Graft type (BM only vs. other)	0.113	0.131	0.389	0.119	0.140	0.395
Karnofsky score (per 10 points)	−0.080	0.071	0.264	−0.085	0.077	0.273
Death in remission:						
Disease (AML vs. ALL)	0.294	0.106	0.006	0.344	0.122	0.005
Age (per 10 years)	0.251	0.036	<0.001	0.284	0.041	<0.001
Graft type (BM only vs.'other)	0.028	0.106	0.795	0.028	0.121	0.815
Karnofsky score (per 10 points)	−0.163	0.052	0.002	−0.183	0.061	0.003

Table 7 Fine–Gray regression models for the competing risks cumulative incidences of relapse and death in remission. The models were also fitted to the pseudo-observations computed at all event times

	Fine–Gray			pseudo-obs. (all t)		
Covariate	β	SE	p	β	SE	p
Relapse:						
Disease (AML vs. ALL)	0.453	0.131	<0.001	0.434	0.134	0.001
Age (per 10 years)	−0.097	0.046	0.036	−0.096	0.047	0.041
Graft type (BM only vs. other)	0.100	0.125	0.424	0.114	0.131	0.387
Karnofsky score (per 10 points)	−0.040	0.069	0.563	−0.084	0.072	0.241
Death in remission:						
Disease (AML vs. ALL)	0.324	0.098	<0.001	0.302	0.105	0.004
Age (per 10 years)	0.262	0.033	<0.001	0.251	0.035	<0.001
Graft type (BM only vs. other)	0.145	0.098	0.142	0.041	0.104	0.695
Karnofsky score (per 10 points)	−0.145	0.050	0.004	−0.161	0.051	0.002

3.4.4 Dependent censoring

The inference relies on an assumption of independent censoring, cf. the discussion above. To check this assumption we fitted a Cox model with censoring as the event, showing that censoring depends on disease and graft type (higher censoring rates for AML and for the BM only graft type). To study the robustness of our results we re-calculated pseudo-observations for the five year survival probability based on the mixture estimator (3) for the four groups given by combinations of these two covariates and fitted the same model as in Table 4 (cloglog link), cf. Table 8. It is seen that the deviations are only minor.

Table 8 Regression model for the 5-year survival probability based on pseudo-observations from the mixture estimator (3)

Covariate	β	SE	p
Disease (AML vs. ALL)	0.399	0.086	<0.001
Age (per 10 years)	0.157	0.029	<0.001
Graft type (BM only vs. other)	0.212	0.085	0.013
Karnofsky score (per 10 points)	−0.125	0.042	0.003

4 Graphical goodness-of-fit based on pseudo-observations

For standard linear regression models, e.g. Draper and Smith,[18] a number of graphical methods are available for checking model assumptions. These include scatterplots and plots based on residuals. In this section we will show how such techniques may be based on pseudo-observations, both for the models studied in the previous sections and for standard models from survival analysis, including the Cox model and the Fine–Gray model. Our presentation will follow that of Pohar Perme and Andersen[10] and illustrations will be based on the bone marrow transplantation data introduced in Section 3.

4.1 Pseudo-residuals

The pseudo-observations are defined for each individual (and, for some examples, at each time point) and can therefore be used to construct residuals analogous to the residuals in a general linear model. The outcome, in our case $\widehat{\theta}_i(t)$, is compared to the predicted value for this outcome based on the model. We denote the predicted value by $\widehat{\theta}(t|Z_i)$ and a raw residual can, therefore, be defined as $\widehat{\theta}_i(t) - \widehat{\theta}(t|Z_i)$.

We will use standardised residuals where the raw residual is divided by a standard error. For example, for the survival indicator, we will divide by an estimate of what would be the standard error of $\widehat{S}_i(t)$ without censoring, that is:

$$\frac{\widehat{S}_i(t) - \widehat{S}(t|Z_i)}{\sqrt{\widehat{S}(t|Z_i)[1 - \widehat{S}(t|Z_i)]}}.$$

We shall use these residuals as a graphical diagnostic tool to evaluate a model fit. In both the Cox model and the Fine–Gray model, two assumptions are made. The coefficient β is assumed to be constant in time (the proportional hazards assumption in the Cox model) and the covariate effect is assumed to be linear. The latter assumption is present in all the models considered.

If the model fits the data well, no trends should be seen in the residuals when we plot them against a covariate (or the linear predictor) at any point of time. On the other hand, when the assumptions of the model are not met, we would like to get some insight into the type of departures. We can expect a nonlinear effect of the covariate to result in a certain trend when plotting the residuals against that covariate. On the other hand, the effect of changing β in time (that is, non-proportional hazards) should be seen in

changes of this trend from time point to time point. In practice, we plot the residuals against the covariate at only a few chosen time points. To make it possible to detect the trends, we shall add a curve representing the smooth average through the residuals.

4.2 Illustration

4.2.1 Survival probabilities, the Cox model

Table 4 shows the estimates in the cloglog (and logit) model for the 5-year survival probability. In Figure 8 the standardised pseudo-residuals from this model are plotted against age (in decades) and Karnofsky. Based on the smoother there seems to be some tendency that residuals are positive for large and for small values of age and negative in between. On the other hand, we see positive residuals for low values of Karnofsky. However, it should be kept in mind that there are few patients with small values of Karnofsky and with either low or high ages, cf. Figure 9.

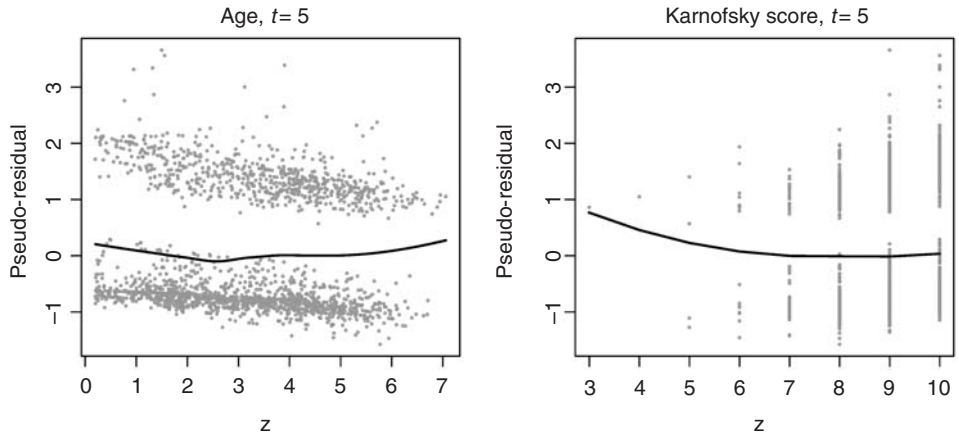

Figure 8 Goodness-of-fit for the $S(5)$ cloglog regression based on pseudo-residuals.

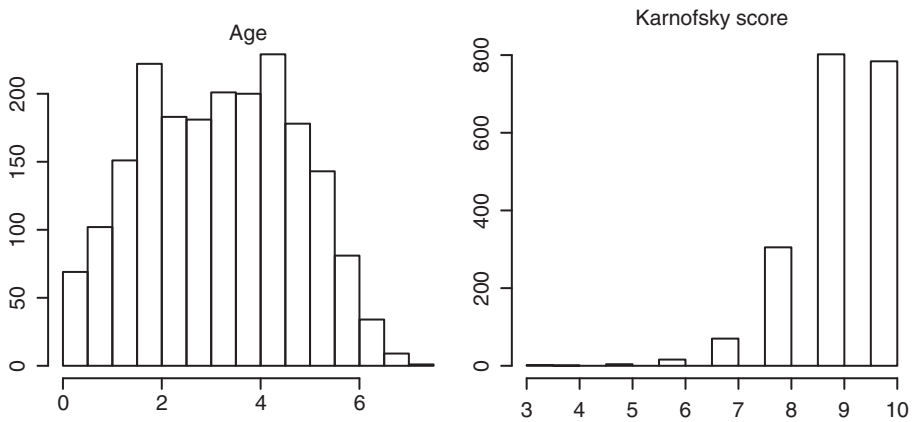

Figure 9 Age and Karnofsky score distribution.

Table 9 Cox model for BMT data

	β	SE	p
Disease (AML vs. ALL)	0.389	0.08	<0.001
Age (per 10 years)	0.154	0.027	<0.001
Graft type (BM only vs. other)	0.161	0.08	0.043
Karnofsky score (per 10 points)	−0.132	0.04	0.001

We also fitted a standard Cox model for the overall mortality in the BMT study, see Table 9. The estimated coefficients are quite close to those obtained in the model for $S(5)$, however, with slightly smaller standard errors.

Figure 10 shows plots of the pseudo-residuals against each of the four covariates in the model at time points 0.25, 0.61, 1.33, 5. For the two binary covariates, disease and graft type, each panel contains only two average residual values which have been connected with a straight line. For disease, both averages seem to be close to 0 indicating no lack-of-fit while, for graft type, there may be a slight tendency that the average value at 1 (BM only) is negative in the beginning and increases for larger values of time. For age, the smooth curve seems to vary around 0 with no systematic trends while that for Karnofsky is clearly negative for the (rare) small values of the covariate and more close to 0 for the more common, larger values.

4.2.2 Competing risks cumulative incidence function

For both of the competing risks, relapse and death in remission, pseudo-residuals from the Fine–Gray model were computed at the values of time (0.23, 0.5, 1.21, 5) for which the model for pseudo-observations was fitted, cf. Table 6. Figures 11 and 12 show the plots of pseudo-residuals against the two quantitative covariates, age and Karnofsky. For age, no indication of lack fit is present while, for Karnofsky, residuals for small values of the covariate tend to be positive.

In conclusion, plots for pseudo-residuals are easy to construct and they may give some indication of lack-of-fit. However, because of the scale of the residuals, i.e. the mean value scale rather than that of the linear predictor, it may be difficult to diagnose what causes an apparent lack-of-fit. To this end, scatterplots to be defined in the next subsection may be superior.

4.3 Pseudo-scatterplots

For standard regression models and for a single covariate, Z_{i1}, a scatterplot of the outcome X_i versus Z_{i1} is useful for assessing how $\mathrm{E}(X_i \mid Z_{i1})$ varies with the covariate. The can be done for both quantitative and binary outcomes and, for the latter, superposition of a scatterplot smoother is essential. In order to evaluate a given link function (e.g. the logistic or the cloglog link) the smoother may be transformed using that link to see whether the transformed curve is roughly linear. A similar plot may be applied for pseudo-observations: plot $\widehat{\theta_i}$ against Z_{i1}, smooth and transform the smoothed curve using the link function.

For multiple covariates, however, the situation is more complex. This is because if the model specifies $g(\mathrm{E}(X_i \mid Z_{i1}, \ldots, Z_{ip}))$ to be linear then $g(\mathrm{E}(X_i \mid Z_{i1})$ need not be

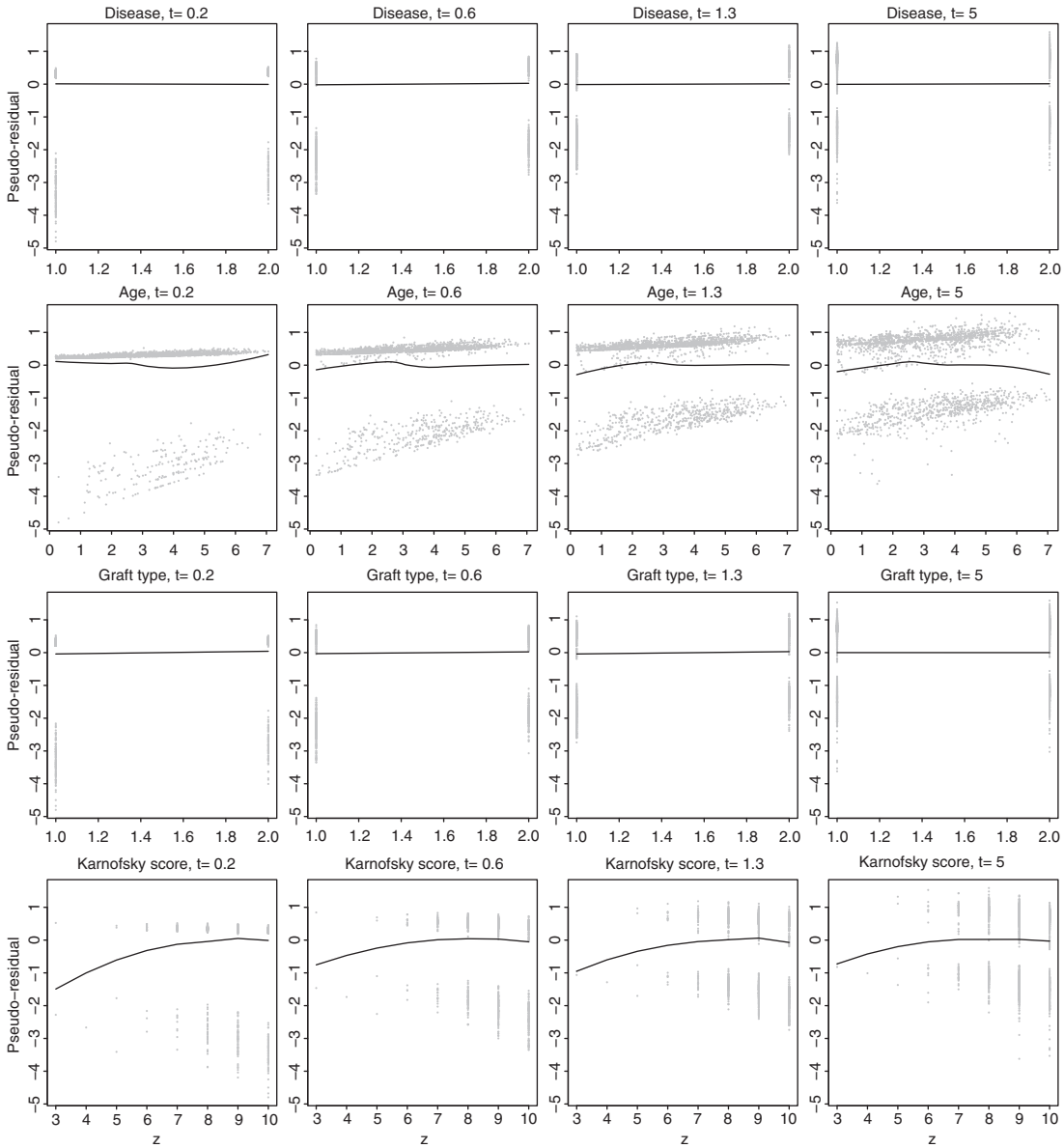

Figure 10 Cox model fit for the model given in Table 9 based on pseudo-residuals. Each row presents one of the variables (disease, age, graft type, Karnofsky score), each column presents one of the four time points chosen (0.25, 0.61, 1.33, 5).

linear (unless g is the identity and covariates are independent). For linear models one then often uses residuals as described in the previous subsection. Here, one may plot the residuals $X_i - \widehat{\mathrm{E}}(X_i \mid Z_{i1}, \ldots, Z_{ip})$ against Z_{i1} to see whether any structure, e.g. due to mismodelling of Z_{i1}, is present. A scatterplot smoother may be added to the plot. A similar smooth plot may be obtained by first doing the scatterplot: X_i versus Z_{i1} and

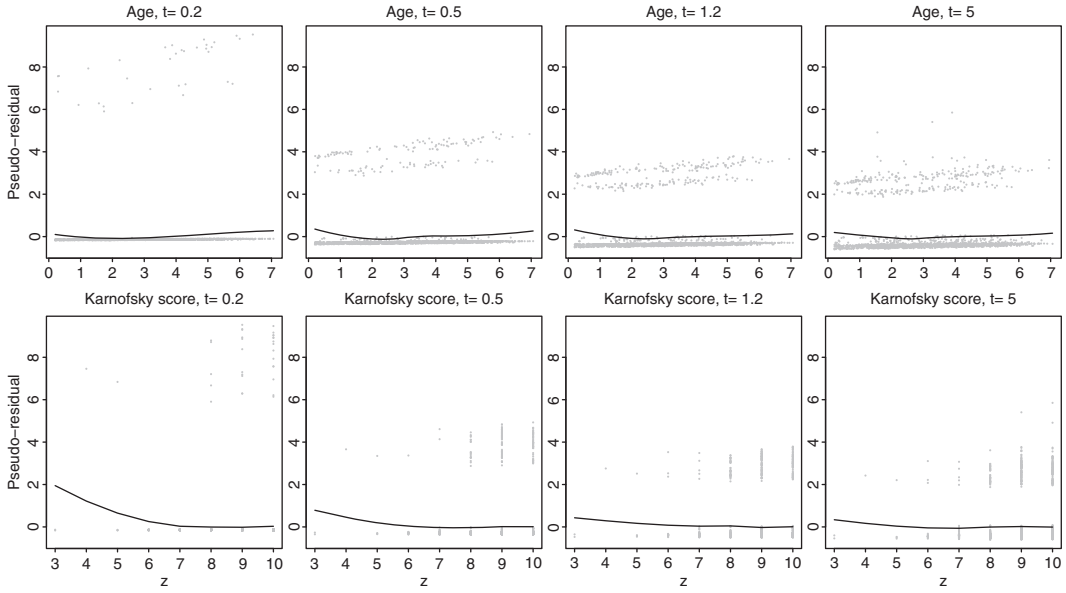

Figure 11 Goodness-of-fit for the competing risks model (relapse) based on pseudo-residuals from the Fine–Gray model. First row: age, second row: Karnofsky. Each column presents one of the four time points: 0.23, 0.5, 1.21. 5.

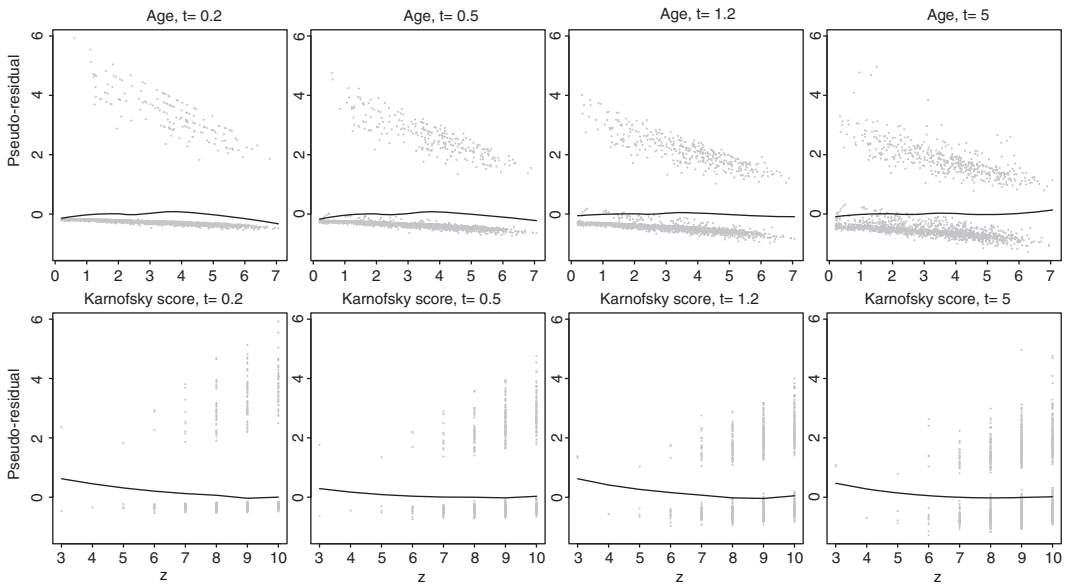

Figure 12 Goodness-of-fit for the competing risks model (death) based on pseudo-residuals from the Fine–Gray model. First row: age, second row: Karnofsky. Each column presents one of the four time points: 0.23, 0.5, 1.21. 5.

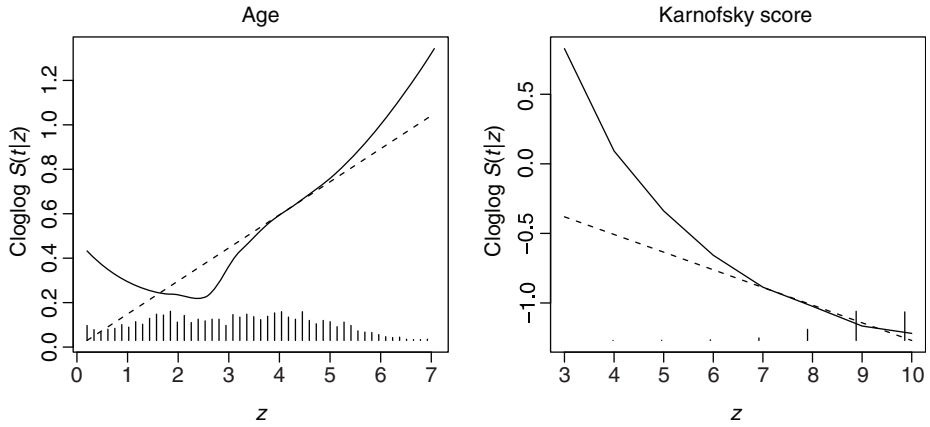

Figure 13 Goodness-of-fit for the $S(5)$ cloglog regression. The straight line represents $\widehat{\beta}Z$.

smooth, next plot $\widehat{\mathrm{E}}(X_i \mid Z_{i1}, \ldots, Z_{ip})$ against Z_{i1} and smooth and, finally, subtract the two smoothers.

This procedure may be adapted to generalised linear models with links other than the identity by transforming the two smoothers using the link function before subtraction. Similarly, the technique may be used for pseudo-observations. For selected values of t (when appropriate):

(1) Plot $\widehat{\theta}_i(t)$ against Z_{i1}, smooth and transform the smoothed curve using the link function. Note that this is simply the one-covariate scatterplot suggested above.
(2) Plot $\widehat{\theta}(t \mid Z_{i1}, \ldots, Z_{ip})$ against Z_{i1}, smooth and transform the smoothed curve using the link function.
(3) Subtract the two smoothed curves. To get a feeling for the magnitude of the variation of this difference, the straight line $\widehat{\beta}_1 Z_{i1}$ may be added.

We will now illustrate this technique for the BMT data. We will supplement the graphs with tests using the GEE as explained in Section 3 to check:

- the linearity of age – by allowing a restricted cubic spline with 3 degrees of freedom (Karnofsky has too few distinct values to do this);
- the change of the coefficient in time (when appropriate) – by allowing separate coefficients at the different time points.

Note that each of these tests, in their simplest form, assumes the other assumption (linearity or time-constant coefficient) to be fulfilled. However, both assumptions may be tested simultaneously by also letting the spline coefficients vary between time points, see Pohar Perme and Andersen[10] for details.

4.4 Illustration

4.4.1 Survival probabilities, the Cox model

Figure 13 shows the goodness-of-fit plots for the cloglog model for the 5-year survival probability. The two quantitative covariates age and Karnofsky are studied and the

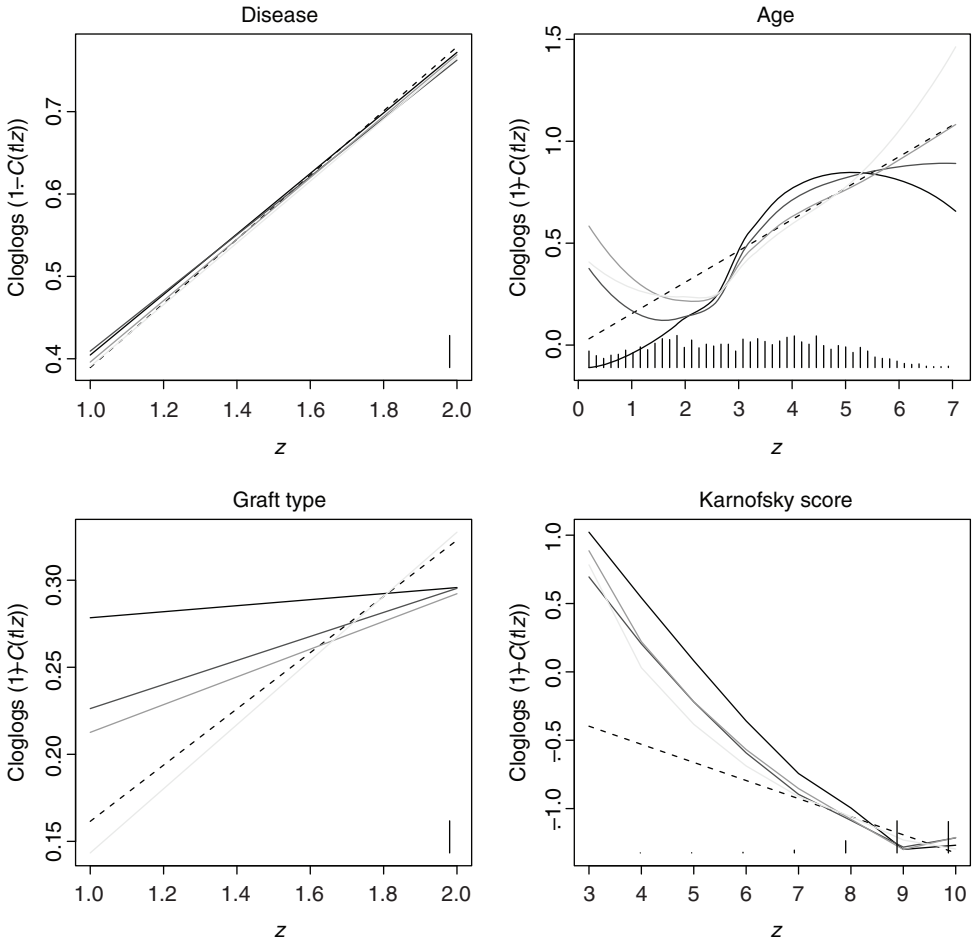

Figure 14 Cox model fit for the model given in Table 9. Each curve represents a different time point (0.25, 0.61, 1.33, 5), the brighter the shade of grey, the later the time. The dashed lines represent $\widehat{\beta}Z$.

straight lines with slopes 0.149 and −0.127 (resp. 0.185 and −0.158) have been added, cf. Table 4. The curve for age is rather close to this line while that for Karnofsky seems to deviate for small values of the covariate. The test for linearity of age confirms the visual impression ($P = 0.12$). However, the corresponding figure for the logit link (not shown) looks very similar and it is not easy to choose between these two link functions.

For the Cox model (Table 9) the corresponding plots are shown in Figure 14 for all of the four covariates in the model. The fits for disease and age seem quite satisfactory while one curve for graft type ($t = 0.25$), in accordance with Figure 10, seems to deviate from those for the later time points. For the bulk of the data (Karnofsky ≥ 60) the curves seem to agree with the line while, for small values, some clear deviations are seen. The numerical tests supplement these observations: the tests for constant effects over time reject for graft type ($P = 0.007$) and Karnofsky ($P = 0.006$, assuming linearity) but not

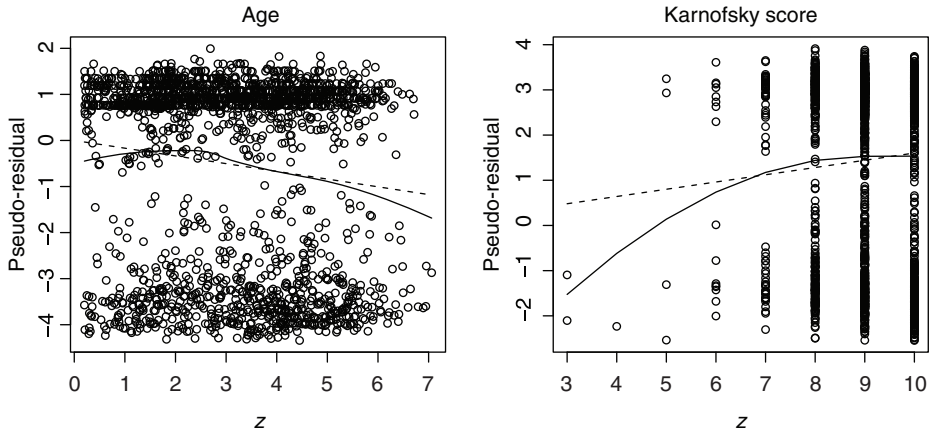

Figure 15 Restricted mean regression: residual plots of pseudo-observations against age and Karnofsky. The dashed lines represent $\widehat{\beta}Z$ that has been added before smoothing.

for age ($P = 0.50$, assuming linearity) or disease ($P = 0.15$). The test for linear effect of age (assuming no changes in time) gives $P = 0.15$ while the overall test for age gives $P = 0.24$.

4.4.2 Restricted mean survival

A linear model was fitted for the 5-year restricted mean, cf. Table 5. Figure 15 shows the plots for the quantitative covariates age and Karnofsky, both seem slightly non-linear. For age, this tendency is confirmed using the numerical test, $P = 0.033$. Since the model is linear, residuals are plotted and the line with slope $\widehat{\beta}$ is added to the smoother.

4.4.3 Competing risks cumulative incidence function

Finally, we turn to the competing risks model. Attention is restricted to the cloglog link and the goodness-of-fit plots for the Fine–Gray model for all covariates are shown in Figures 16 (relapse) and 17 (death). For relapse it should be kept in mind that only disease is clearly significant. For the binary covariates the assumption of constant effect over time is clearly accepted using the numerical tests ($P = 0.19, 0.59$, respectively for disease and graft type). For age and Karnofsky the tests are boundary significant ($P = 0.04, 0.03$, respectively). The test for linearity of age and the overall tests both give $P = 0.07$. For death in remission, the test for time-constant effect of the highly significant variable disease clearly rejects ($P = 0.005$) though the curves look rather similar. For the insignificant graft type, the test gives $P = 0.002$ in accordance with the plot. The effect of Karnofsky, assuming linearity, varies significantly between time points, $P = 0.004$, while the model assumptions seem reasonable for age (all $P > 0.37$).

Figures 16 and 17 present the scatterplots where the predicted values are based on the Fine–Gray model. However, as was the case for the Cox model, similar plots could be

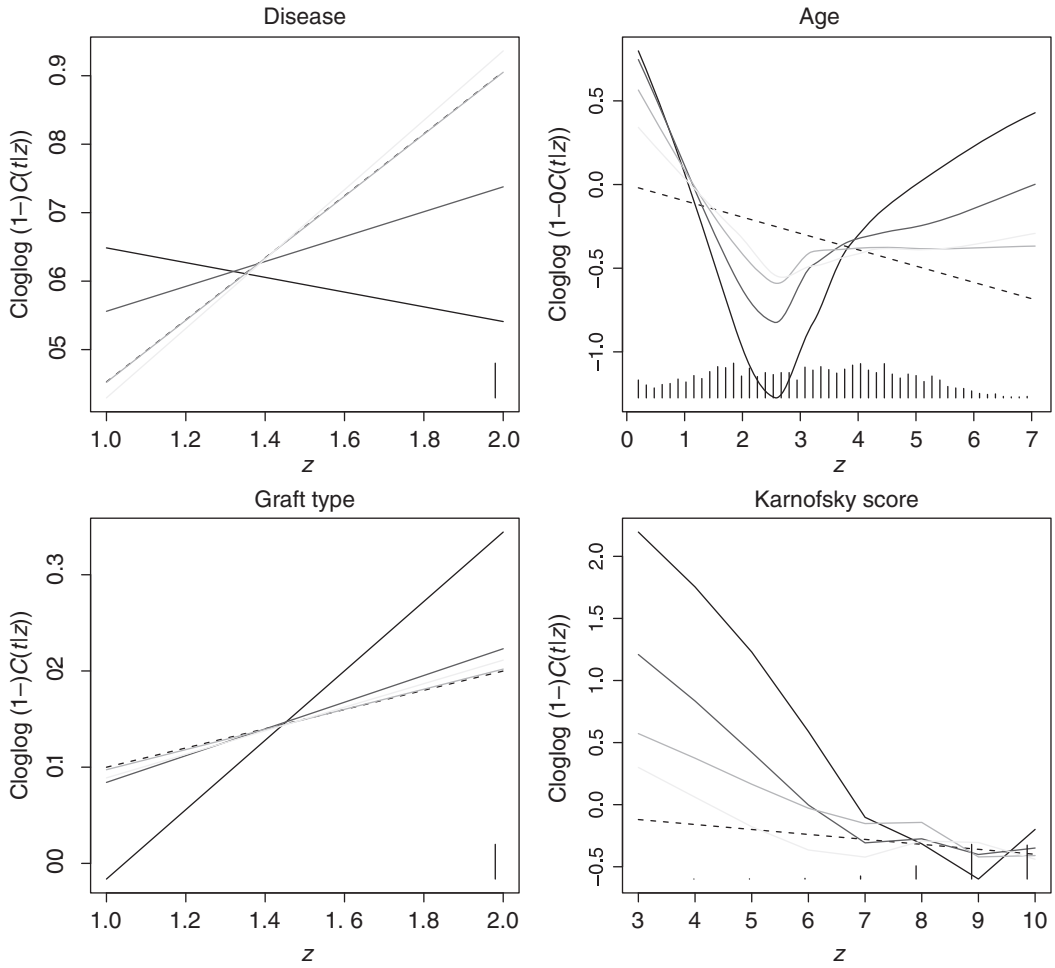

Figure 16 Goodness-of-fit for the Fine–Gray competing risks model – relapse. Each curve represents a different time point (0.23, 0.5, 1.21, 5), the brighter the shade of grey, the later the time. The dashed lines represent $\widehat{\beta}Z$.

based on a model fitted to pseudo-observations. The corresponding figures (not shown) reveal the same tendencies.

5 Discussion

Pseudo-observations were orignally introduced in connection with applications of jackknife methods for, e.g. bias reduction.[19] More recently, pseudo-observations have got a revival in survival and event history analysis with the purpose of doing regression for certain mean value parameters, θ, e.g. transition and state occupation probabilities in multi-state models,[4,7] in particular competing risks cumulative incidences,[6] restricted mean survival time,[5] and the survival function in a single time point.[8]

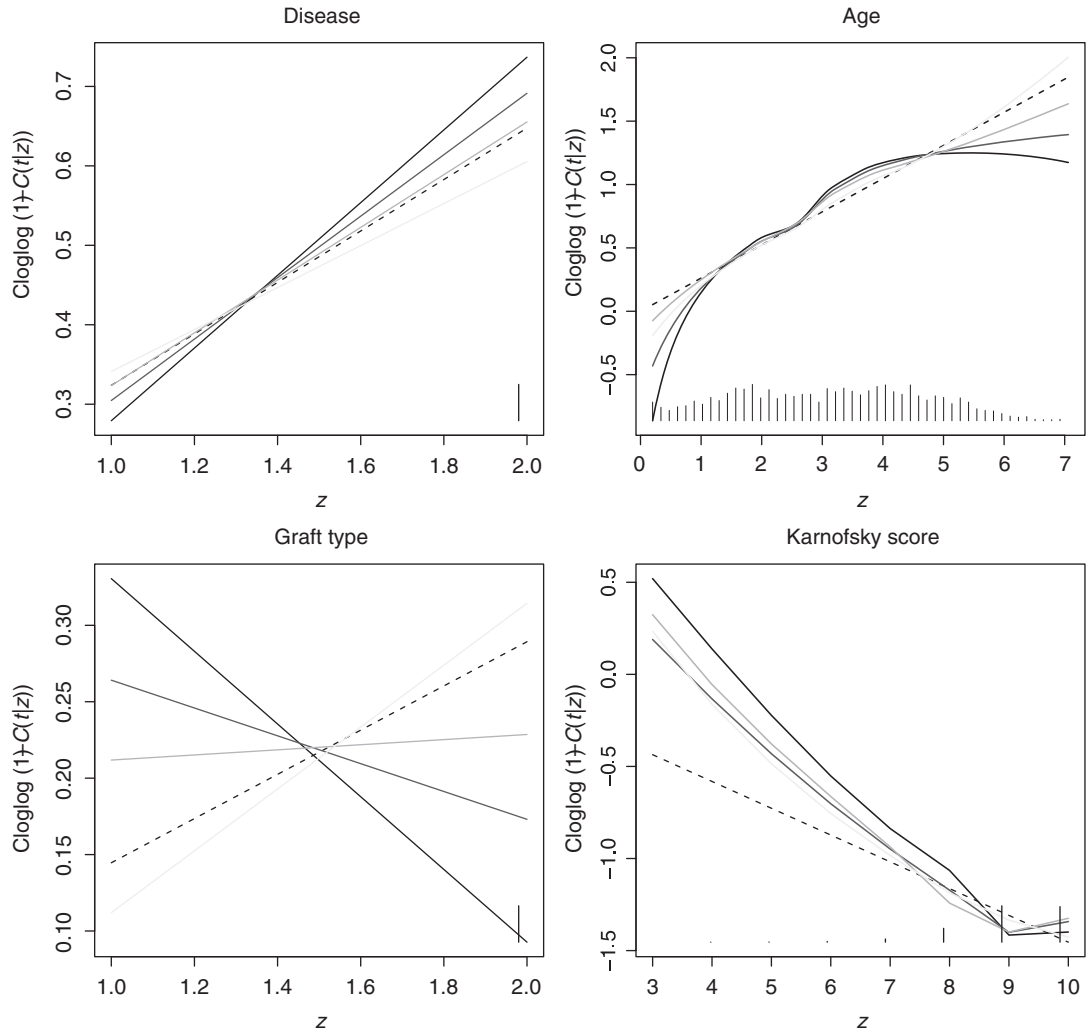

Figure 17 Goodness-of-fit for the Fine–Gray competing risks model – death. Each curve represents a different time point (0.23, 0.5, 1.21, 5), the brighter the shade of grey, the later the time. The dashed lines represent $\widehat{\beta}Z$.

We reviewed this development in Section 3. Further references in this area include Helms *et al.*,[20] who studied insurance premiums, and Andrei and Murray,[21] who analysed quality-of-life-adjusted life times.

With the article by Graw *et al.*[11] the mathematical basis for some of this development has been established. Even though the results of Graw *et al.* only cover the cumulative incidence (and special cases thereof, see Section 2) and only a fixed set of time points at which pseudo-observations are calculated, there is a hope that the techniques they use (e.g. influence functions and compact differentiability) are sufficiently general to allow for extensions to other situations.

A requirement for using pseudo-observations is that the estimator for θ is unbiased. Therefore, for the typical survival data situation, censoring should not depend on covariates since otherwise, e.g. the Kaplan–Meier estimator $\widehat{S}(t)$ is not unbiased for the marginal survival distribution function $\theta = S(t)$. We investigated this problem in a small simulation study and came up with an alternative mixture estimator for $S(t)$ when censoring depends on discrete covariates. More work, however, is needed on consequences of dependent censoring when using pseudo-observations.

When using pseudo-observations for model fitting, a choice of time points needs to be made and though simulations suggest that the sensitivity of the results to this choice is not great, it would be nice to avoid the choice altogether and to use 'all' time points. In Section 3, we did a step in this direction and estimated regression effects for the cumulative incidence function based on all time points, however, with a smoothness restriction for the baseline function.

A more recent development in the application of pseudo-observations in survival analysis is the article by Pohar Perme and Andersen,[10] surveyed in Section 4, dealing with graphical (and numerical) goodness-of-fit examination. This method provides, among other things, techniques for examining model assumptions for hazard regression models like the Cox model or an additive model,[22] and for the Fine and Gray regression model for cumulative incidences. Though pseudo-observations may also be used for estimation in these models this is not what we recommend since classical estimators are more efficient. More experience with the goodness-of-fit plots obtained in this way is needed and, in particular, it would be useful to supplement the graphs with confidence limits. We are currently studying how the bootstrap may be used for that purpose.

Finally, it should be mentioned that although we have focused on the use of pseudo-observations in survival and event history analysis there may be other areas of application, e.g. regression models for correlation coefficients or for variance parameters in linear models may be obtained in this way.

The pseudo-observations for the examples presented in this article can be calculated using the `pseudo` package in `R`,[9] the graphs can be plotted by means of the `R` function `pseucheck`, available at:

`http://www.mf.uni-lj.si/ibmi-english/biostat-center/index.html`

Acknowledgements

The research was supported by grant R01-54706-12 from the National Cancer Institute and Danish Natural Science Research Council, grant 272-06-0442 'Point process modelling and statistical inference'. We are grateful to CIBMTR for providing us with the example data.

References

1 Kalbfleisch JD, Prentice R. *The statistical analysis of failure time data* (2nd edn). John Wiley and Sons, Inc. Hoboken, NJ; 2002.

2 Andersen PK, Borgan Ø, Gill RD, Keiding N. *Statistical models based on counting processes*. Springer-Verlag, New York; 1993.

3 Parmar M, Machin D. *Survival analysis: A practical approach*. John Wiley and Sons, Inc., Chichester; 1995.
4 Andersen PK, Klein JP, Rosthøj S. Generalised linear models for correlated pseudo-observations, with applications to multi-state models. *Biometrika* 2003; **90**: 15–27.
5 Andersen PK, Hansen MG, Klein JP. Regression analysis of restricted mean survival time based on pseudo-observations. *Lifetime Data Analysis* 2004; **10**: 335–50.
6 Klein JP, Andersen PK. Regression modeling of competing risks data based on pseudovalues of the cumulative incidence function. *Biometrics* 2005; **61**: 223–9.
7 Andersen PK, Klein JP. Regression analysis for multistate models based on a pseudo-value approach, with applications to bone marrow transplantation studies. *Scandinavian Journal of Statistics* 2007; **34**: 3–16.
8 Klein JP, Logan BL, Harhoff MG, Andersen PK. Analysing survival curves at a fixed point in time. *Statistics in Medicine* 2007; **26**: 4505–19.
9 Klein JP, Gerster M, Andersen PK, Tarima S, Pohar Perme M. SAS and R functions to compute pseudo-values for censored data regression. *Computer Methods and Programs in Biomedicine* 2008; **89**: 289–300.
10 Pohar Perme M, Andersen PK. Checking hazard regression models using pseudo-observations. *Statistics in Medicine* 2008; **27**: 5309–28.
11 Graw F, Gerds TA, Schumacher M. On pseudo-values for regression analysis in multi-state models. *Lifetime Data Analysis* 2009; **15**: 241–255.
12 Andersen, P. K. Keiding, N. Multi-state models for event history analysis. *Statistical Methods in Medical Research* 2002; **11**: 91–115.
13 Tukey JW. Bias and confidence in not quite so large samples. *Annals of Mathematical Statistics* 1958; **29**: 614.
14 Stute W, Wang JL. The jackknife estimate of a Kaplan-Meier integral. *Biometrika* 1994; **81**: 603–6.
15 Fine JP, Gray RJ. A proportional hazards model for the subdistribution of a competing risk. *Journal of the American Statistical Association* 1999; **94**: 496–509.
16 Liang KY, Zeger SL. Longitudinal data analysis using generalised linear models. *Biometrika* 1986; **78**: 13–22.
17 Fine JP, Yan J, Kosorok MR. Temporal process regression. *Biometrika* 2004; **91**: 683–703.
18 Draper NR, Smith H. *Applied regression analysis* (3rd edn). John Wiley and Sons, Inc., New York; 1998.
19 Miller RG. The jackknife – a review. *Biometrika* 1974; **61**: 1–15.
20 Helms F, Czado C, Gschlössl S. Calculation of LTC premiums based on direct estimates of transition probabilities. *ASTIN Bulletin* 2005; **35**: 445–69.
21 Andrei A-C, Murray S. Regression models for the mean of the quality-of-life-adjusted restricted survival time using pseudo-observations. *Biometrics* 2007; **63**: 398–404.
22 Lin DY, Ying Z. Semiparametric analysis of the additive risk model. *Biometrika* 1994; **81**: 61–71.