Medical University of Vienna
Center for Medical Statistics, Informatics and Intelligent Systems
Section for Clinical Biometrics
A-1090 VIENNA, Spitalgasse 23

Phone: (+43)(1) 40400/66880
Fax: (+43)(1) 40400/66870
http://cemsiis.meduniwien.ac.at/kb

Technical Report 06/2014

%PSHREG: A SAS® Macro for Proportional and Nonproportional Substribution Hazards Regression for Survival Analyses with Competing risks

Maria Kohl and Georg Heinze

e-mail: georg.heinze@meduniwien.ac.at

Abstract

We present a new SAS macro %PSHREG that can be used to fit a proportional subdistribution hazards (PSH) model (Fine and Gray, 1999) for survival data subject to competing risks. Our macro first modifies the input data set appropriately and then applies SAS's standard Cox regression procedure, PROC PHREG, using weights and counting-process style of specifying survival times to the modified data set (Geskus, 2011). The modified data set can also be used to estimate cumulative incidence curves for the event of interest. The application of PROC PHREG has several advantages, e.g., it directly enables the user to apply the Firth correction, which has been proposed as a solution to the problem of undefined (infinite) maximum likelihood estimates in Cox regression, frequently encountered in small sample analyses (Heinze and Schemper, 2001).

In case of non-PSH, the PSH model is misspecified, but offers a time-averaged summary estimate of the effect of a covariate on the subdistribution hazard (Grambauer, Schumacher and Beyersmann, 2010). Random censoring usually distorts this summary estimate compared to its expected value had censoring not occured, as later event times are underrepresented due to earlier censorship. The solution would be upweighting late event times in the estimating equations by the inverse probability of being observed, similarly to Xu and O'Quigley's (2000) proposal for reweighting the summands of the estimating equations in the Cox model. A very appealing interpretation of the average subdistribution hazard ratio as odds of concordance can be obtained by weighting the summands by the expected number of patients at risk (Schemper, Wakounig and Heinze, 2009). Both types of weights are available in %PSHREG. We illustrate application of these extended methods for competing risks regression using our macro, which is freely available at http://cemsiis.meduniwien.ac.at/en/kb/science-research/software/statistical-software/PSHREG, by means of analysis of real and artificial data sets.

Contents

1 Overview

Competing risks arise in the analysis of time-to-event data, if the event of interest is impossible to observe due to a different type of event occuring before. Competing risks may be encountered, e.g., if interest focuses on a specific cause of death, or if time to a non-fatal event such as stroke or myocardial infarction is studied. In both situations, death from a non-disease-related cause would constitute the competing risk.

It has frequently been pointed out that in presence of competing risks, the standard product-limit method of describing the distribution of time-to-event yields biased results. The main assumption of this method is that any subject whose survival time is censored will experience the event of interest if followed up long enough. This does not hold if competing risks are present, as the occurence of the event of interest is made impossible by an antecedent competing event. As a remedy, the cumulative incidence estimate proposed by Kalbfleisch and Prentice (1980)can be used. While the product-limit estimate of cumulative probability of an event will reach 1 with an infinite follow-up time, the cumulative incidence estimate never reaches 1 as a consequence of presence of a certain proportion of subjects who will experience the competing event.

Two different ways of modeling competing risks data have been proposed. The first one analyses the cause-specific hazard of each event type separately, by applying Cox regression targeting each event type in turn, and censoring all other event types. By a complete analysis of all event types, estimated cumulative incidence curves for an event of interest can be estimated.

By contrast, the proportional subdistribution model proposed by Fine and Gray (1999) directly aims at modeling differences in the cumulative incidence of an event of interest. Its estimation is based on modified risk sets, where subjects experiencing the competing event are retained even after their event. In case of censoring (which is the rule rather than the exception), a modification of this simple principle was proposed such that the weight of those subjects artificially retained in the risk sets is gradually reduced according to the conditional probability of being under follow-up had the competing event not occurred.

A SAS–macro `%PSHREG` was written to implement the model proposed by Fine and Gray (1999), in SAS (SAS Institute Inc., 2010). The macro modifies an input data set by separating follow-up periods of patients with competing events into several sub-periods with declining weights, following the suggestions in Fine and Gray (1999) and Geskus (2011). This allows to use SAS's PROC PHREG to compute the proportional subdistribution hazards model. PROC PHREG can either be called automatically by the macro, or the user may call it after the macro with the modified data set. This possibility allows to make full use of the various options that PROC PHREG offers for modeling and output control.

All options offered by PROC PHREG for verifying and relaxing the assumption of proportional subdistribution hazards can be used, including the computation and display of unscaled and scaled Schoenfeld-type residuals. Similarly to the implementation in R (library `cmprsk`), the `%PSHREG` macro can deal with time-dependent effects of covariates to accommodate non-proportional subdistribution hazards, allowing to specify interactions of covariates with any functions of time. In addition, non-proportionality may also be accounted for by computing weighted estimates that are connected to the odds of concordance as defined by Schemper et al. (2009) for standard survival analyses. For completeness, also inverse-probability-of-censoring weights can be applied, as suggested for Cox regression by Xu and O'Quigley (2000).

In very small data sets with few events, monotone pseudo-likelihood may cause parameter estimates to diverge to $\pm\infty$. This phenomenon usually happens if events are observed in only one of two levels of a binary covariate. In this case, the robust standard error will collapse to zero, while the model-based standard error diverges with the parameter estimate. The application of the Firth-correction,

which is readily implemented in PROC PHREG, may be useful in such circumstances. It penalizes the likelihood such that parameter estimates are optimally corrected for small-sample bias, and always leads to finite estimates (Heinze and Schemper, 2001). For confidence interval computation, the profile penalized likelihood can be used, which is valid because the reweighting of the data set does not concern the event times (Geskus, 2011).

In the remainder of this report we first briefly review the estimation of proportional subdistribution hazards models with time-fixed and time-dependent effects, and introduce weighted estimation in the proportional subdistribution hazards model. The third section explains all the macro options in detail. The report closes with a worked example using a publicly available data set.

2 Methods

2.1 An example for competing risks

We consider a study on lymphoma presented in Pintilie (2006). In this study, researchers were interested in estimating the effect of risk factors on time to a relapse of lymphoma, which may also affect the overall survival time. Death constitutes a competing event, if occurring before the relapse. The data set includes 541 patients from the Princess Margaret Hospital of Toronto with an follicular type lymphoma diagnosed between 1967 and 1996. All patients were at an early stage of disease (stage I or stage II) and had been treated with radiation alone (RT) or with radiation and chemotherapy (CMT). Also the age and the haemoglobin value of the patients are known. The event of interest is time from diagnosis until relapse or 'no response' on treatment. Age, haemoglobin value, clinical stage and treatment are the risk factors of interest and are considered in modeling the cumulative incidence of relapse.

2.2 Cause-specific cumulative incidence estimation

Without loss of generality, we assume that there is one event type of interest (index 1) and only one competing event (index 2). Hence, in our example, relapse is event type 1 and death before relapse event type 2. Let $h_k(t)$ and $H_k(t)$ denote the cause-specific hazard function and cause-specific cumulative hazard functions, respectively, for cause k, $k = 1, 2$. The cause-specific cumulative incidence function $F_k(t)$, describing the cumulative proportion of subjects experiencing event type 1 up to time t, is given by

$$F_1(t) = \int_0^t S(s)h_1(s)ds \tag{2·1}$$

Note that $S(t)$ is the survival function of time to first of the two event types, given by $S(t) = e^{-H_1(t)-H_2(t)}$. $F_k(t)$ has also been denoted as the 'subdistribution', reflecting the fact that it does not reach 1 in presence of a competing risk. Figure 1 displays the relapse-specific cumulative incidence function over time separately for patients with clinical stage 1 and clinical stage 2.

In the absence of competing risks, the cumulative hazard and cumulative incidence (one minus survivor) function are connected by the relationship $F(t) = 1 - e^{-H(t)}$. This unique correspondence is lost with competing risks, because the cumulative incidence for the event of interest depends on the cause-specific hazard of the competing event (Andersen et al., 2012). Consequently, the Kaplan-Meier estimator of the cumulative incidence function is biased and the inequality $1-S_j(t) \geq F_j(t)$, with $j = 1, 2$, holds (Bakoyannis and Touloumi, 2012). The Kaplan-Meier estimator would estimate the cumulative incidence function in the situation where the competing risk could be eliminated and its elimination would not change the cause-specific hazard. Expressed differently, the Kaplan-Meier estimator pretends that the competing risk, similarly as censoring, is a feature of the study at hand which will not occur in the target population out there.

The cumulative incidence function $F_1(t)$ at the event times t_i, $i = 1, \ldots, m$ can be estimated by

$$\hat{F}_1(t_i) = \sum_{s=t_1}^{t_i} d_{1i}/n_i \hat{S}(t_{i-1}) \tag{2·2}$$

where d_{1i} is the number of events of type 1 observed at t_i, n_i is the number of patients at risk just before t_i, and $\hat{S}(t_{i-1})$ is the Kaplan-Meier estimator of the survival function of time to first event.

An alternative estimate is based on the empirical cumulative subdistribution hazard function estimate and presented later.

2.3 Cause specific analysis

The standard approach to relate the time to the event of interest to covariates is to model the cause-specific hazard semiparametrically, using a Cox regression model with competing events treated like censored observations. However, prediction of the cumulative incidence function from a cause-specific hazards regression model is not straightforward, since estimators from cumulative cause-specific hazards for both the event of interest and the competing event are needed.

In the cause-specific analysis, separate models are fit for each case of events, corresponding to the proportional hazards model with the time to the first event that occurs. These models provide estimates of the effect of variables on the cause-specific hazard, but not on the cumulative incidence of events, since cumulative hazard and cumulative incidence are not connected (see above). Cause specific analysis provides relative measures of the effect of a variable on the risk of the event of interest. Cause-specific hazards regression is directly available in PROC PHREG. However, for cumulative incidence function estimation following cause-specific hazards regression, specialised software is needed, such as the SAS macro of Rosthøj et al. (2004).

A cause-specific analysis censors subjects at the time at which a competing event is observed. Thus, the results apply to the population actually at risk for the event of interest at each time, irrespective of observed rates of competing events.

2.4 Proportional subdistribution hazard regression

2.4.1 Introduction

Proportional subdistribution hazards regression analysis evaluates effects of covariates on the 'subdistribution hazard' which is the basis of the cause-specific cumulative incidence function. In its basic representation, it assumes that the effects of covariates on the subdistribution hazard are stable over time. Patients who experience a competing event are left 'forever' in the risk set (but with decreasing weight to account for declining observability). Consequently, the results apply to populations with similar rates of competing events as the sample at hand (Pintilie, 2007).

2.4.2 Estimation of model parameters

We consider T as the time at which the first event of any type occurs in an individual, and ϵ the event type related to that time. The subdistribution hazard $\gamma(t, X)$ is defined as

$$\gamma(t, X) = \lim_{\Delta t \to 0} \frac{1}{\Delta t} \Pr\{t \leq T \leq t + \Delta t, \epsilon = 1 | T \geq t \cup (T \leq t \cap \epsilon \neq 1), X\} \tag{2·3}$$

with X denoting a row vector of covariates.

Following Fine and Gray (1999), it can be modeled as a function of a parameter vector β through

$$\gamma(t, X) = \gamma_0(t) e^{X\beta} \tag{2·4}$$

where γ_0 is the baseline hazard of the subdistribution.

The partial likelihood of the subdistribution hazards model was defined by Fine and Gray (1999) as

$$L(\beta) = \prod_{j=1}^{r} \frac{\exp(x_j \beta)}{\sum_{i \in R_j} w_{ji} \exp(x_i \beta)} \tag{2·5}$$

where r is the number of all time points ($t_1 < t_2 < ... < t_r$) where an event of type 1 occured, and x_j is the covariate row vector of the subject experiencing an event of type 1 at t_j. For simplicity, here no ties

in event times are assumed. The weights w_{ji} are needed as soon as censoring occurs. The risk set R_j is defined as

$$R_j = \{i;\ t_i \geq t\ \cup\ (t_i \leq t\ \cap \epsilon_i \neq 1)\} \tag{2·6}$$

At each time point t_j, the set of individuals at risk R_j includes those who are still at risk of that event type as well as those who have had a competing event before time point t_j. Subjects without any event of interest prior to t_j participate fully in the partial likelihood with the weight $w_{ji} = 1$, whereas time-dependent weights are defined for subjects with competing events prior to t_j, as

$$w_{ji} = \frac{\hat{G}(t_j)}{\hat{G}(\min(t_j, t_i))} \tag{2·7}$$

Here, $\hat{G}(t)$ denotes the product-limit estimator of the survival function of the censoring distribution, i.e., the cumulative probability of still being followed-up at t. These latter individuals have weights $w_{ji} \leq 1$, which decrease with time.

The proportional subdistribution hazards model can be estimated using any standard software for Cox regression that allows for counting process representation of times (start-stop syntax) and weighting (Geskus, 2011). This is accomplished by using unmodified data on the subjects who either experience event type 1 or who are censored, and modifying only the observations on the subjects who experience event type 2.

In particular, each survival time t_i is represented in counting process style as one or several conjoint episodes. For individuals with event type 1 or censored times, these episodes, denoted by (start time, stop time, status indicator) are just $(0, t_i, d_i^*)$, where the modified censoring indiciator d_i^* is 1 for event type 1, and 0 for a censored time. However, observations on subjects experiencing event type 2 are modified. Here, the first episode is given by $(0, \max_{t_l<t_i}(t_l), 0)$, and to reflect the artificial retaining of those individuals in the risk sets, the following episodes $(t_j, t_{j+1}, 0)$, $t_j \geq \max_{t_l<t_i}(t_l)$ are generated for all following event times until t_r. These episodes are assigned the decreasing weights w_{ji}.

2.4.3 Prediction of cumulative incidence

Let $Y_i(t) = I(t_i > t)$, and $N_i(t) = 1$ if $T_i \geq t \cup d_i* = 1$, and $N_i(t) = 0$ otherwise. Thus, $dN_i(t)$ is the increment in the counting process describing the status of subject i with respect to event type 1 in the interval $[t, t + dt)$. (This counting process changes from 0 to 1 at the event time T_i if the event type 1 has occurred at that time.) The baseline cumulative subdistribution hazard, relating to an individual with a zero covariate vector, is given by

$$\hat{\Lambda}_{10}(t) = \sum_{i=1}^{n} \int_0^t \frac{dN_i(s)}{\sum_{j=1}^n w_{ji} Y_j(s) \exp(x_j \hat{\beta})} \tag{2·8}$$

With time-invariant covariates X, the empirical cumulative distribution hazard for event type 1 is given by $\hat{\Lambda}_1(t, X) = \exp(X\hat{\beta})\hat{\Lambda}_{10}(t)$. The empirical cumulative subdistribution hazard estimate of the cumulative incidence function can then simply be estimated by $\hat{F}(t, X) = 1 - \exp\{-\hat{\Lambda}_1(t, X)\}$.

2.4.4 Monotone likelihood and Firth's bias correction method

In fitting a Cox regression model, the phenomenon of monotone likelihood is observed if the likelihood converges while at least one entry of the parameter estimate diverges (Heinze and Schemper, 2001). The same may happen in a PSH model, e.g., if events of interest are only observed in one of two levels of a binary explanatory variable.

In case of monotone likelihood, not only the parameter estimate but also its standard error diverges. Thus, inference based on standard errors becomes uninformative, even if based on values observed at the last iteration before the likelihood converged. On the other hand, the robust standard error as proposed by Lin and Wei (1989), which is also used by default in Fine-Gray models, collapses to zero in case of monotone likelihood. This standard error is based on DFBETA residuals $\delta_i = \beta - \beta^{(i)}$, which are one-step approximations to the Jackknife values (Therneau and Grambsch, 2000), i.e., the changes in parameter estimates if each observation in turn is left out from analysis. (Specifically, with $\boldsymbol{\delta} = \delta_1, \ldots, \delta_n$, the robust variance matrix V is computed as $V = \boldsymbol{\delta}'\boldsymbol{\delta}$.) Omitting observations from the estimation process does not help in making the parameter estimate converge, since $\beta \to \infty$ implies $\beta^{(i)} \to \infty$, $i = 1, \ldots, n$. Thus, the robust standard error of a divergent parameter estimate will collapse to 0.

By adding an asymptotically negligible penalty function to the log likelihood, the occurrence of divergent parameter estimates can be completely avoided (Firth, 1993; Heinze and Schemper, 2001). Furthermore, the penalty, suggested for exponential family models in canonical representation by Firth (1993) as $1/2 \log |I(\beta)|$, with $I(\cdot)$ denoting the Fisher information matrix, corrects the small sample bias of maximum likelihood estimates. This bias is usually low for Cox regression models unless monotone likelihood is observed. Estimation can be based on modified score functions, and in PSH models the only additional feature are the weights that go into the modified estimation procedure.

In case of monotone likelihood, it was proposed that inference should be based on the profile penalized likelihood function, since the normal approximation may fail because of the asymmetry of the profile penalized likelihood (Heinze and Schemper, 2001). With $\ell(\boldsymbol{\beta})$ denoting the log of the likelihood, and $\ell_{\max}$ its maximum value, the profile log likelihood function of parameter β_j is given by $\ell_j^*(\gamma) = \max_{\boldsymbol{\beta}\setminus\beta_j} \ell(\boldsymbol{\beta}|\beta_j = \gamma)$, i.e., by the log likelihood fixed at $\beta_j = \gamma$ and maximized over all parameters except β_j. $2(\ell_{\max} - \ell_j^*)$ has a limiting χ^2 distribution with 1 degree of freedom. Let $\ell_0 = \ell_{\max} - 1/2\chi_1^2(1-\alpha)$. Thus, a $(1-\alpha) \times 95\%$ confidence interval for β_j can be obtained by $\{\gamma : \ell_j^*(\gamma) \geq \ell_0\}$. Profile penalized likelihood confidence intervals are simply obtained by exchanging $\ell(\boldsymbol{\beta})$, $\ell_{\max}$ and $\ell(\boldsymbol{\beta}|\beta_j = \gamma)$ by their penalized versions.

2.5 Nonproportional subdistribution hazards

2.5.1 Schoenfeld-type residuals

Similarly as in the Cox PH model, as a first explorative step, Schoenfeld-type residuals and weighted Schoenfeld-type residuals can be inspected in order to detect violations of the proportional subdistribution hazards assumption. At the event time t_j of the i^{th} subject having covariate row vector X_i, a row vector of Schoenfeld-type residuals is defined by $\hat{U}_i(t_j) = X_i) - \bar{X}(\boldsymbol{\beta}, t)$, where $S^{(0)}(\boldsymbol{\beta}, t) = \sum_i Y_i(t) w_{ji} \exp(X_i\boldsymbol{\beta})$, $S^{(1)}(\boldsymbol{\beta}, t) = \sum_i Y_i(t) X_i w_{ji} \exp(X_i\boldsymbol{\beta})$, and $\bar{X}(\boldsymbol{\beta}, t) = S^{(1)}(\boldsymbol{\beta}, t)/S^{(0)}(\boldsymbol{\beta}, t)$. Weighted Schoenfeld-type residuals are scaled such that the smoothed residuals can directly interpreted as changes in β over time. They are defined as $\boldsymbol{r}_i = n_{e1} I^{-1}(\hat{\boldsymbol{\beta}}) \hat{\boldsymbol{U}}_i(t_j)$, with n_{e1} denoting the number of events of type 1.

2.5.2 Time-varying coefficients

The proportional subdistribution hazards model lends itself to accommodate non-proportional hazards of covariates by including time-varying covariates defined by products of covariates with functions of time. The basic model is extended in the following way:

$$\gamma(t, x) = \gamma_0(t) e^{X\beta(t)} \tag{2·9}$$

with $\beta(t) = f(t, \beta)$. Considering a single covariate, then, in its simplest form, $f(t, \beta)$ could be defined as $\beta_1 + \beta_2 t$, such that a covariate's effect is modeled as increasing or decreasing linearly with time. To allow for complex dependencies, flexible functions of time such as splines (Durrleman and Simon, 1989) or fractional polynomials ($\beta_1 + \beta_2 t^{p_1} + \beta_2 t^{p_2}$), with p_1, p_2 selected from a pre-defined set of values (Royston and Altman, 1994), could be used.

2.5.3 Population-averaged coefficients

Estimation of an average subdistribution hazard ratio (ASHR) as proposed for Cox regression by Schemper, Wakounig and Heinze (2009) can be obtained by weighting the risk sets in the estimating equations by the expected numbers of subjects at risk, which are defined by $v_j = \{1 - F_1(t_j)\}G^{-1}(t_j)$ with $F_1(t)$ and $G^{-1}(t)$ denoting the cumulative incidence function of the event of interest and the inverse survival function of the censoring distribution, respectively. These weights are multiplied with w_{ji}, such that the weight of individual i in risk set $R(t_j)$ is $v_j \times w_{ji}$.

2.5.4 Time-averaged coefficients

If an effect of a covariate on the subdistribution hazard is not constant over time, then the PSH model is misspecified, if the time-dependency is not accounted for by including appropriate time-dependent terms. The PSH model parameter estimate of such a variable can be seen as a summary estimate (Grambauer et al., 2010). However, the summary estimate may depend on the actual censoring distribution. To make the summary estimate independent of the actual censoring distribution, inverse probability of censoring weights (IPCW) can be applied multiplicative to the Fine-Gray weights, such that the final weight of individual i in risk set $R(t_j)$ is $G^{-1}(t_j) \times w_{ji}$. According to Xu and O'Quigley (2000), this estimates a time-averaged regression effect.

3 Working with the macro

3.1 Syntax

The following options are available in `%PSHREG` (the brackets < and > denote options that need not to be specified):

`%PSHREG(<data=`*SAS data set*`,>`
`time=`*variable*`,`
`cens=`*variable*`,`
`<failcode=`*value*`,>`
`<cencode=`*value*`,>`
`<varlist=`*variables*`,>`
`<class=`*variables*`,>`
`<cengroup=`*variable*`,>`
`<firth=`*value*`,>`
`<options=`*string*`,>`
`<id=`*variable*`,>`
`<action=`*string*`,>`
`<cuminc=`*value*`,>`
`<by=`*variable*`,>`
`<censcrr=`*variable*`,>`
`<out=`*SAS data set*`,>`
`<weights=`*value*`,>`
`<call=`*SAS data set*`,>`
`<missing=`*string*`,>`
`<delwork=`*value*`,>`
`<tiedcens=`*sring*`,>`
`<clean=`*value*`>);`

These options are described in the subsequent sections.

3.2 Basic options

- `data=`*SAS data set* names the input SAS data set. The default value is `_LAST_`.
- `time=`*variable* names a variable containing survival times. There is no default value.
- `cens=`*variable* names a variable containing the censoring indicator for each survival time. There is no default value.
- `failcode=`*value* names the event value. The default value is 1, meaning that if the variable specified in the `cens` option assumes that the value 1, then the corresponding survival time is treated as event.
- `cencode=`*value* names the censoring value. The default value is 0, meaning that if the variable specified in the `cens` option assumes that the value 0, then the corresponding survival time is treated as censored.

- `varlist=`*variables* names a list of independent variables, separated by blanks. There is no default value. This option is required.
- `class=`*variables* names categorical variables, all must also be specified in `varlist`. There is no default value. This option will automatically generate a CLASS statement in the PROC PHREG call and has no other purpose.
- `cengroup=`*variable* Optional: variable with different values for each group with a distinct censoring distribution (the censoring distribution is estimated separately within these groups). This parameter has the same meaning as the cengroup option in the R program `crr`.
- `id=`*variable* may serve as patient identifier. There is no default value.
- `by=`*variable* may define subsets for efficient processing of multiple data sets of the same structure. There is no default value.
- `censcrr=`*variable* defines a new variable in the output data set, which contains the modified status indicator. The default name of this variable is `_censcrr_`.
- `missing=`*string* specifies if missing values in the modified data set should be carried forward to the analysis or the output data (`missing=keep`, default) or if lines with missing values in any variable in the `varlist` option should be deleted (`missing=drop`).
- `delwork=`*value* specifies if all working data sets should be deleted on exit (`delwork=1`, default) or kept (`delwork=0`).
- `tiedcens=`*string* specifies if censored times that are tied with event times should be handled after (`tiedcens=after`, the default) or before (`tiedcens=before`) event times.
- `clean=`*value* if set to 1, requires that the output data set should be cleaned, i.e., keeping only relevant variables mentioned in the macro call

3.3 Weighting options

- `weights=`*value* applies weights to the risk sets in addition to the Fine-Gray weighting. These weights are IPCW weights to estimate a time-averaged effect if `weights=1`, or ASHR weights to estimate a population-averaged effect (odds of concordance-type effect) if `weights=2`.

3.4 Output options

- `out=`*SAS data set* names the output data set including all covariables, the start and stop times of the counting-processes and, if requested by the `weights` option, weights of the observations. The default name is `dat_crr`.
- `action=`*string* requests the estimation of the Fine-Gray proportional subdistribution hazards model using as covariates all variables specified in `varlist` (`action=estimate`, default). If `action=code` PROC PHREG is not invoked, but the code needed to estimate the PSH model via PROC PHREG is printed in the Log window.
- `cuminc=`*value* plots the cumulative incidence curves (stratified by the levels of the first variable specified in `varlist`) if set to 1. The default value is 0.

3.5 Model fitting options

- `firth=`*value* turns the Firth penalization on (`firth=1`) or off (`firth=0`), which solves the phenomenon of monotone likelihood and shrinks the coefficient estimators towards zero.
- `options=`*string* specifies model fitting options which are used by PROC PHREG. For possible values see the documentation of `PROC PHREG`.

3.6 Printed output

In any case, the macro will create a modified data set suitable to estimate a Fine-Gray model using PROC PHREG. Since the weights needed to estimate a Fine-Gray model do not depend on covariates, it is not necessary to repeat this data-modifying step every time a Fine-Gray model should be estimated with different variables. Thus, we have implemented an option which controls whether the Fine-Gray model should be estimated immediately, or if only the modified data set should be created. If `action=estimate`, the macro will estimate the Fine-Gray model. If `action=code` the model will not be estimated, but then the SAS Log window will contain a NOTE with SAS code, which could be submitted (perhaps after modifying it by specifying a different set of explanatory variables etc.) to have the Fine-Gray model computed.

The macro can also compute cumulative incidence curves stratified for the levels of the first variable in `varlist`. Here, the same strategy was applied: if `cuminc=1`, then cumulative incidence curves will be shown, otherwise, the SAS statements needed to obtain these curves will be shown in the SAS Log window.

3.6.1 Output of PSHREG

The first page of output will always contain a list of the macro option values. If `action=estimate`, then additional pages will contain the results from the Fine-Gray model.

3.6.2 SAS code generation

If `action=code`, then SAS code will be written into the SAS Log window. This SAS code can be copied to the Editor and submitted to estimate the Fine-Gray model. Some researchers may want to use different sets of variables, or transformations or interactions of variables. It is not necessary to repeat the macro call in this case; once the modified data set is created, the user can apply PROC PHREG with different variable lists etc. in the same manner as shown in the example code of the SAS Log.

3.7 Computational issues

`%PSHREG` does not do any statistical computations besides calling PROC LIFETEST to compute survival probabilities in order to compute the time-dependent weights. All statistical computation is passed over to PROC PHREG, which employs well-validated algorithms to estimate the models. All parameters to control the iterative estimation procedure offered by PROC PHREG (convergence criteria, ridging, etc.) can be used.

4 Examples

4.1 A macro call using default settings

The use of %PSHREG is exemplified using the aforementioned data set of the follicular non-Hodgkin lymphoma study. The data set is available at: http://www.uhnres.utoronto.ca/labs/hill/datasets/Pintilie/datasets/follic.txt (15 October 2013) and can be read into SAS by the statements

```
filename rawfoll URL
 'http://www.uhnres.utoronto.ca/labs/hill/datasets/Pintilie/datasets/follic.txt';


data follic;
infile rawfoll firstobs=2 delimiter="," DSD;
input age path1 $ hgb ldh clinstg blktxcat  relsite $ ch $ rt $ survtime stat dftime
      dfcens resp $ stnum;
run;
```

Time from diagnosis until relapse is coded in a variable named dftime. We would like to model time to relapse, taking into account the competing risk of death without relapse. To this end, we generate an event status variable evcens and a competing risk status variable crcens following the description in Pintilie (2006) by the code below. For applying the macro, we need another status variable combining the information in evcens and crcens using the levels 0, 1 and 2. The patients' ages, their haemoglobin values, their clinical stages and their treatments (chemotherapy or other) are used as explanatory variables.
In particular, we submit the following data step statements:

```
data follic;
  set follic;
  if resp='NR' or relsite^='' then evcens=1; else evcens=0;
  if resp='CR' and relsite='' and stat=1 then crcens=1; else crcens=0;
  cens=evcens+2*crcens;
  agedecade=age/10;
  if ch='Y' then chemo=1; else chemo=0;
run;
```

A proportional subdistribution hazards model, using default settings of the macro options, is estimated by submitting:

```
%phsreg(data=follic, time=dftime, cens=cens, varlist=agedecade hgb clinstg chemo);
```

Results are illustrated below. The output first shows a page with the selected macro options, and then includes a summary of the number of events, competing events and censored values. The remainder of the output is produced by PROC PHREG. Note that the number of observations given here refers to the number of distinct lines in the modified data set and is usually much greater than the number of subjects.

The PSHREG macro: summary of macro options

Macro option	Assigned value	Remark
data	follic	Input data set
time	dftime	Time variable
cens	cens	Censoring variable
failcode	1	Code for event of interest
cencode	0	Code for censored observation
tiedcens	after	How censored times tied with event times should be treated
varlist	agedecade	List of covariables
	hgb	
	clinstg	
	chemo	
class		List of class variables
options		Options to be passed to PROC PHREG
firth	0	Standard ML estimation, no Firth correction
id		Subject identifier
by		BY processing variable
cuminc	0	Requests cumulative incidence curves
action	estimate	Fine-Gray model computed.
weigths	0	Standard model, no weighting of risk sets
clean	1	Unnecessary variables removed
call	_PSHREGopt	Data set with this call‘s macro options
out	dat_crr	Output data set for standard Fine-Gray mode
missing	keep	Keep lines with missing covariate values
statustab	1	Summary of status variable requested
delwork	1	Temporary data sets deleted on exit
macro version	2014.06	
build	201406250855	

The PSHREG macro: Summary of status variable

Obs	_status	COUNT	PERCENT
1	Censored	193	35.6747
2	Events of interest	272	50.2773
3	Competing events	76	14.0481

The PSHREG macro: Fine-Gray model

The PHREG Procedure

Model Information

Data Set	WORK.DAT_CRR
Dependent Variable	_start_
Dependent Variable	_stop_
Censoring Variable	_censcrr_
Censoring Value(s)	0
Weight Variable	_weight_
Ties Handling	BRESLOW

Number of Observations Read	9875
Number of Observations Used	9799

Summary of the Number of Event and Censored Values

Total	Event	Censored	Percent Censored
9799	272	9527	97.22

Convergence Status

Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics

Criterion	Without Covariates	With Covariates
-2 LOG L	3198.496	3170.556
AIC	3198.496	3178.556
SBC	3198.496	3192.979

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	27.9400	4	<.0001
Score (Model-Based)	27.7283	4	<.0001
Score (Sandwich)	23.7832	4	<.0001
Wald (Model-Based)	27.5975	4	<.0001
Wald (Sandwich)	24.8896	4	<.0001

Analysis of Maximum Likelihood Estimates

Parameter	DF	Parameter Estimate	Standard Error	StdErr Ratio	Chi-Square	Pr > ChiSq	Hazard Ratio
agedecade	1	0.17251	0.04794	1.044	12.9494	0.0003	1.188
hgb	1	0.00231	0.00398	0.995	0.3377	0.5611	1.002
clinstg	1	0.55658	0.13507	1.021	16.9809	<.0001	1.745
chemo	1	-0.33198	0.17295	1.040	3.6846	0.0549	0.718

Two variables have a significant influence on the cumulative incidence of relapse: age and clinical stage.

While the macro's main purpose is to compute such proportional subdistribution hazards models, it may also be used to estimate (unadjusted) cumulative incidence curves. Assume we would like to estimate cumulative incidence curves according to clinical stages. The following macro call can be used:

```
%PSHREG(data=follic, time=dftime, cens=cens, varlist=clinstg, cuminc=0, action=code);
```

`cuminc` could be set to 1 to directly plot the cumulative incidence curves. Setting `cuminc=0`, however, gives the user more flexibility with respect to graphical parameters etc. The option `action=code` precludes the estimation of the Fine-Gray model. Since in the above call `cuminc=0`, the following code is created in the Log window:

```
proc phreg data=dat_crr ;
    model (_start_,_stop_)*_censcrr_(0)=;
    weight _weight_;
```

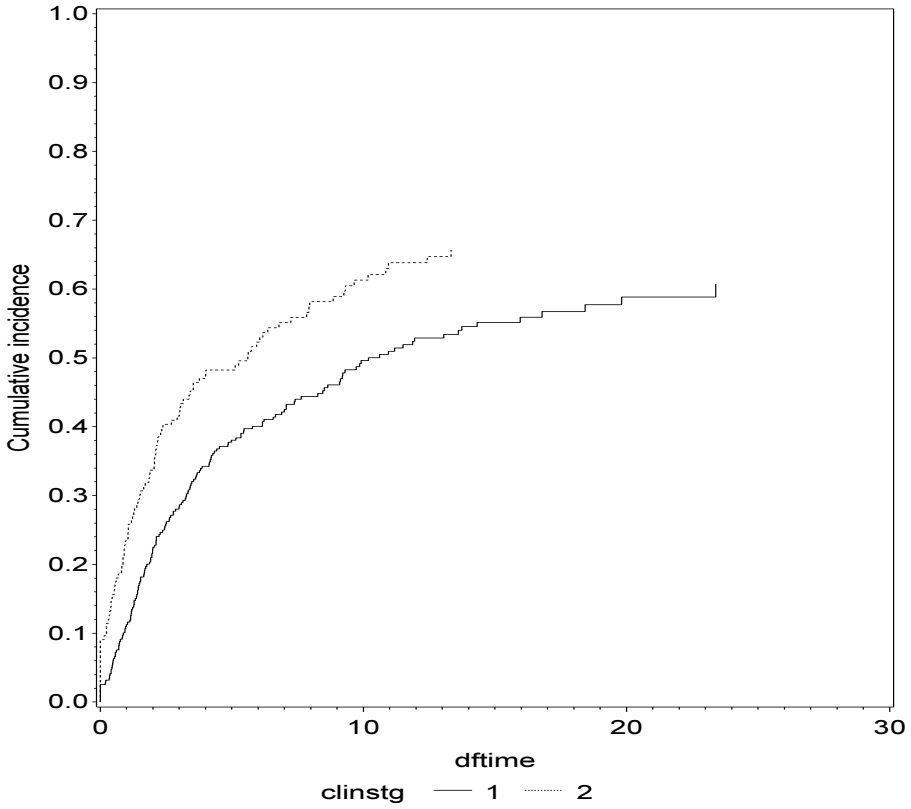
```
    strata clinstg;
    baselin out=_cuminc survival=_surv /method=EMP;
run;

data _cuminc;
    set _cuminc;
    _cuminc_=1-_surv;
    dftime=_stop_;
    label _cuminc_="Cumulative incidence";
    drop _stop_ _surv;
run;

symbol1 I=steplj LINE=1 C=black;
symbol2 I=steplj LINE=2 C=black;

proc gplot data=_cuminc;
    plot _cuminc_*dftime=clinstg;
run;
```

By copying this code chunk from the Log window into the program Editor window, graphical parameters can easily be redefined by the user, such as line or symbol colors, labeling, etc, again directly using all features offered by SAS. This offers optimal flexibility in presenting results. Also the estimation method of the cumulative incidence curves can be modified, default setting is `method=EMP` (see also documentation of PROC PHREG). Figure 1 show the cumulative incidence curves resulting from the SAS code above.

Alternatively, the cumulative incidence curve could be plotted with the SAS-supplied `CUMINCID` macro:

```
%cumincid(data=follic, time=dftime, status=cens, event=1, compete=2, censored=0, strata=clinstg)
```

Figure 1: Cumulative incidence of relapse by clinical stage I vs. II, estimated with the PSHREG macro.



4.2 Scaled Schoenfeld-type residuals

For detection of time-dependent effects it is useful to evaluate Schoenfeld residuals. The following code describes how Schoenfeld residuals and a restricted cubic spline can be plotted in SAS. In this example we assume that Schoenfeld residuals of the variables `age, hgb, clinstg` and `chemo` should be computed and plotted (see Figures 2, 3, 4 and 5).
We start with the macro call

```
%PSHREG(data=follic, time=dftime, cens=cens, varlist=clinstg agedecade hgb chemo,
        action=code)
```

which will not estimate the Fine-Gray model, but will generate some SAS code in the Log window. This code can then be copied into the Editor window and modified in the following way:

```
proc phreg data=dat_crr covs(aggregate) out=estimates;
    model (_start_,_stop_)*_censcrr_(0)=agedecade hgb clinstg chemo;
    output out=schoenfeld_data  wtressch=WSR_agedecade WSR_hgb WSR_clinstg WSR_chemo;
    id _id_;
    weight _weight_;
    by _by_;
run;
```

The third line of the code chunk above (the `output` statement) creates a new data set, `schoenfeld_data`, containing the weighted Schoenfeldtype residuals for all variables in the model and for all event time

points. Submission of the modified code will compute the Fine-Gray model, and the code will also generate two new data sets, `estimates` (containing parameter estimates) and `schoenfeld_data` (containing the Schoenfeld-type residuals).

In following code chunk we merge these two data sets. For the data merger, it is necessary to specify a key variable. We can make use of the constant `_by_`, which is automatically generated by the macro (if the `by` option was not used) and which assumes the value of 1 for all lines in `schoenfeld_data` as well as in `estimates`.

```
data schoenfeld_data;
merge schoenfeld_data(keep=dftime _by_ WSR_agedecade WSR_hgb
   WSR_clinstg WSR_chemo) estimates;
by _by_;
rescaled_WSR_agedecade=WSR_agedecade+agedecade;
rescaled_WSR_hgb=WSR_hgb+hgb;
rescaled_WSR_clinstg=WSR_clinstg+clinstg;
rescaled_WSR_chemo=WSR_chemo+chemo;
ldftime=log(dftime+1);
label rescaled_WSR_agedecade="beta(t) of age per decade"
      rescaled_WSR_hgb="beta(t) of haemoglobin"
      rescaled_WSR_clinstg="beta(t) of stage"
      rescaled_WSR_chemo="beta(t) of chemotherapy"
      ldftime="log of time";
run;
```

In the data step above, we rescale the residuals by adding the parameter estimates. Rescaled and smoothed residuals have the interpretation of time-dependent parameter estimates. Smoothing can be performed using `PROC LOESS` as described below, and by making use of `ods graphics` the raw and smoothed time-dependent parameters along with their 95% confidence limits can be displayed (code only shown for `agedecade`):

```
ods graphics on;
ods select fitplot;
proc loess data=schoenfeld_data plots=residuals(smooth);
    model rescaled_WSR_agedecade=ldftime /CLM smooth=0.5;
run;
ods graphics off;
```

Figure 2: Schoenfeld residuals for age.

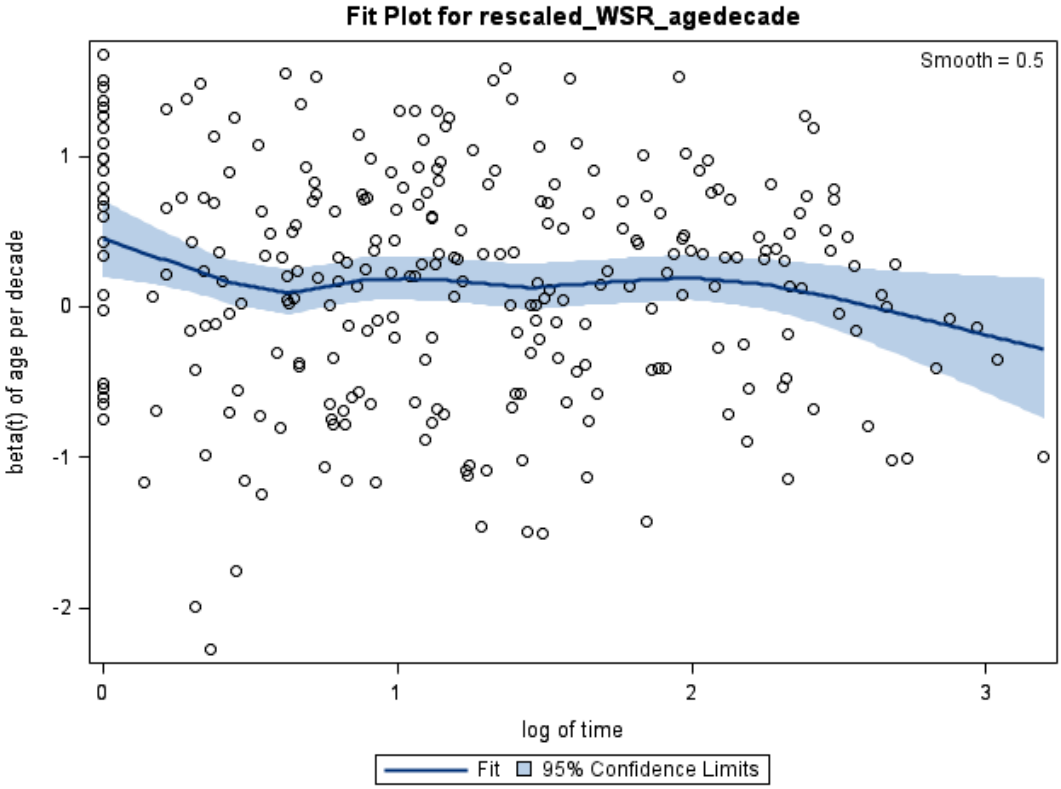


Figure 3: Schoenfeld residuals for haemoglobin.

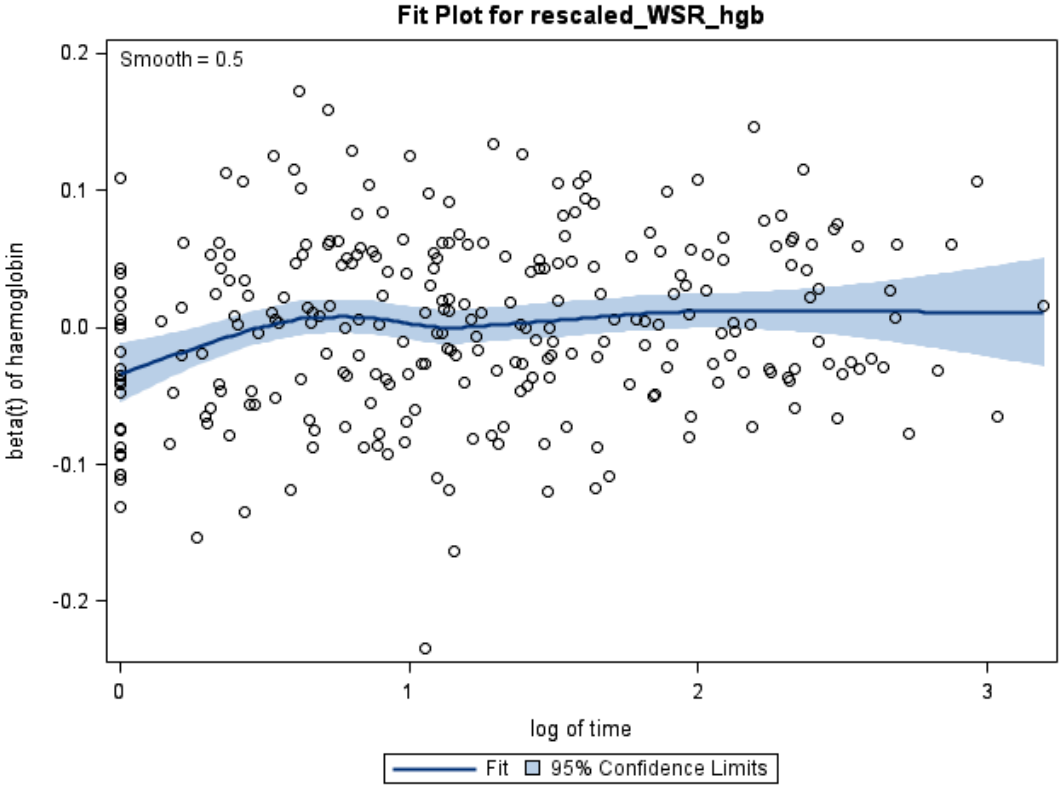


Figure 4: Schoenfeld residuals for clinical stage.

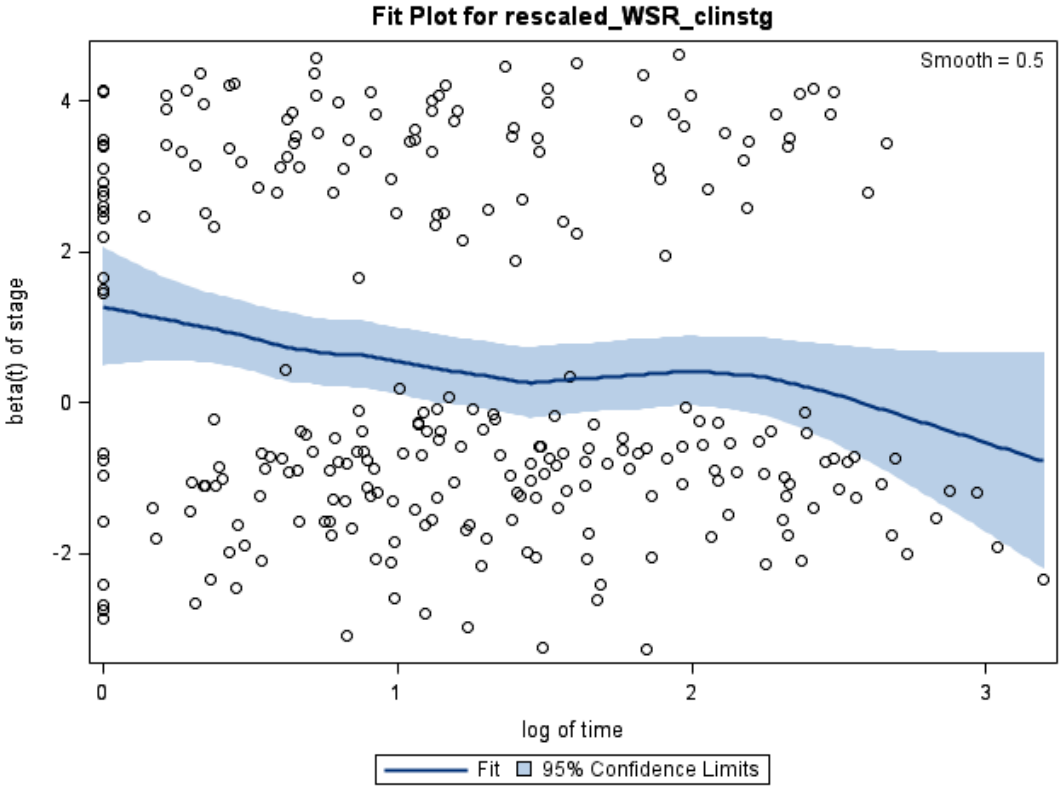
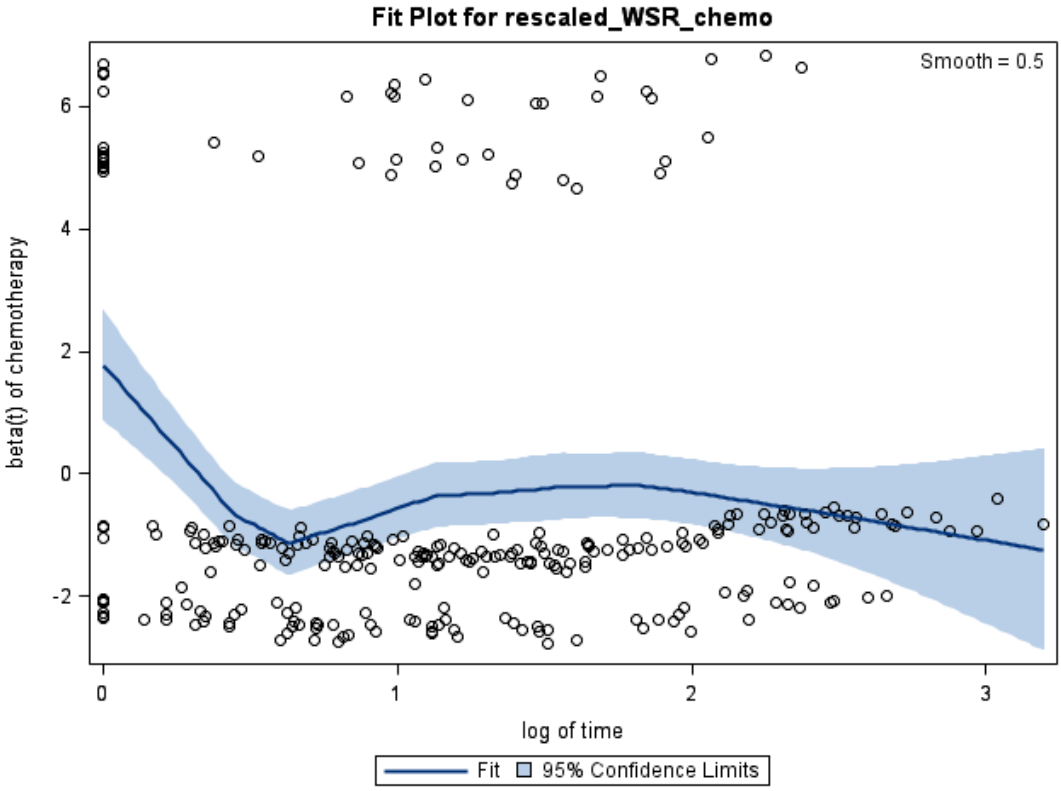


Figure 5: Schoenfeld residuals for chemotherapy.



4.3 Time-varying coefficients

As it is obvious from the Schoenfeld residuals plot, the variable `clingstg` (clinical stage) shows a time dependent effect (see Figure 4). To estimate its time-dependent effect on the subdistribution hazard, i.e., to relax the proportional subdistribution hazards assumption, we specify the following statements:

```
proc phreg covs(aggregate) data=dat_crr ;
model (_start_,_stop_)*_censcrr_(0)=agedecade hgb clinstg clinstg*logstop1 chemo;
    logstop1=log(_stop_+1);
    id _id_;
    weight _weight_;
    hazardratio clinstg/at(logstop1=0 1.79 2.40) ;
run;
```

A working variable `clinstg*logstop1` is defined, which defines the kind of time-dependency of the effect of `clinstg`. Here, we define `logstop1` as the logarithm of the time plus one. For an intuitive description of the results it makes sense to show hazard ratios at different time points. Here we specify 0, 5 and 10 years, re-expressed in log(months+1).

The PHREG Procedure

Model Information

Data Set	WORK.DAT_CRR
Dependent Variable	_start_
Dependent Variable	_stop_
Censoring Variable	_censcrr_
Censoring Value(s)	0
Weight Variable	_weight_
Ties Handling	BRESLOW

Number of Observations Read	9875
Number of Observations Used	9799

Summary of the Number of Event and Censored Values

Total	Event	Censored	Percent Censored
9799	272	9527	97.22

Convergence Status

Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics

Criterion	Without Covariates	With Covariates
-2 LOG L	3198.496	3162.645
AIC	3198.496	3172.645
SBC	3198.496	3190.674

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq

Likelihood Ratio	35.8509	5	<.0001
Score (Model-Based)	36.4702	5	<.0001
Score (Sandwich)	29.5061	5	<.0001
Wald (Model-Based)	35.5468	5	<.0001
Wald (Sandwich)	34.0091	5	<.0001

Analysis of Maximum Likelihood Estimates

Parameter	DF	Parameter Estimate	Standard Error	StdErr Ratio	Chi-Square	Pr > ChiSq	Hazard Ratio
agedecade	1	0.16515	0.04754	1.036	12.0660	0.0005	1.180
hgb	1	0.00198	0.00392	0.981	0.2553	0.6134	1.002
clinstg	1	1.10088	0.23152	0.985	22.6098	<.0001	.
logstop1*clinstg	1	-0.46681	0.16781	0.988	7.7383	0.0054	.
chemo	1	-0.32982	0.16994	1.021	3.7669	0.0523	0.719

Hazard Ratios for clinstg

Description	Point Estimate	95\% Wald Confidence Limits	Robust
clinstg Unit=1 At logstop1=0	3.007	1.910	4.733
clinstg Unit=1 At logstop1=1.79	1.304	0.928	1.832
clinstg Unit=1 At logstop1=2.4	0.981	0.599	1.606

Revealed by the output above, there exists a time-dependent effect of clinical stage. A strong effect of stage can only be confirmed for time point zero, at later time points the effect declines.

4.4 Time-averaged and population-averaged analysis

Unlike any other implementation of the Fine-Gray model, the `PSHREG` macro can compute and apply weight functions for the *risk sets* to obtain weighted estimators of parameters and subdistribution hazard ratios. Two different weighting functions are available. For weights according to the inverse probability of being uncensored, use `weights=1`. For estimation of average subdistribution hazard ratios (Schemper et al., 2009), use `weights=2`:

```
%PSHREG(data=follic, time=dftime, cens=cens, varlist=agedecade hgb clinstg chemo,
        weights=1);
```

```
%PSHREG(data=follic, time=dftime, cens=cens, varlist=agedecade hgb clinstg chemo,
        weights=2);
```

The output of both weighting methods is compared below:

Analysis of Maximum Likelihood Estimates

	Parameter	DF	Parameter Estimate	Standard Error	StdErr Ratio	Chi-Square	Pr > ChiSq	Hazard Ratio
IPCW	agedecade	1	0.14050	0.04953	1.166	8.0470	0.0046	1.151
	hgb	1	0.00341	0.00412	1.091	0.6825	0.4087	1.003
	clinstg	1	0.44595	0.14698	1.168	9.2063	0.0024	1.562
	chemo	1	-0.37229	0.17612	1.091	4.4685	0.0345	0.689
AHR	agedecade	1	0.16486	0.04942	0.978	11.1294	0.0008	1.179
	hgb	1	0.00144	0.00402	0.913	0.1274	0.7211	1.001
	clinstg	1	0.55026	0.13951	0.955	15.5577	<.0001	1.734
	chemo	1	-0.30818	0.17568	0.960	3.0771	0.0794	0.735

The values of these two types of weights can be plotted against time by the following statements, leading to the plot shown in Figure 6:

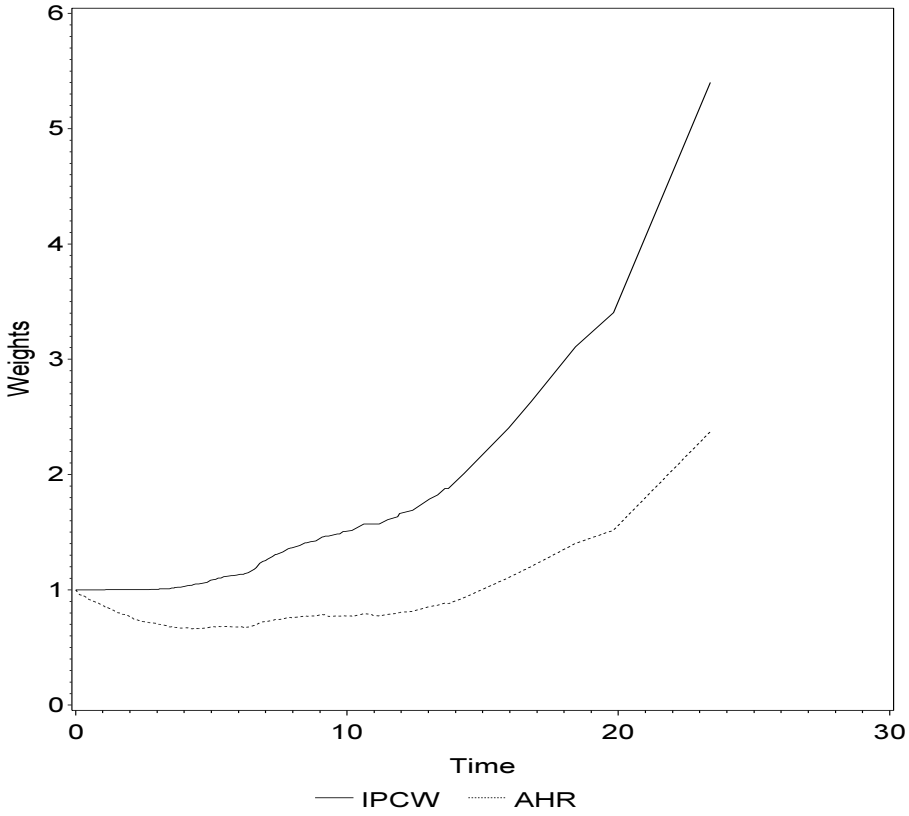
```
symbol1 i=join v=none c=black line=1;
symbol2 i=join v=none c=black line=2;
axis1 label=(angle=90 'Weights');
axis2 label=('Time');
legend1 lable=none value=("IPCW" "AHR");

proc gplot data=dat_crr_w;
    plot (_ipcweight_ _ahrweight_)*_stop_ /overlay vaxis=axis1 haxis=axis2 legend=legend1;
    where _wcens_=1;
run;
```

4.5 Ties handling

PROC PHREG offers various options for handling ties in event times, which all can be adopted by `%PSHREG`. To demonstrate the effect of different ties handling, we have to introduce ties to our data set by rounding the time variable `dftime` to one decimal place:

```
data follicties;
    set follic;
    dftimeties=round(dftime,1);
run;

proc freq data=follicties;
    table dftimeties;
    where cens=1;
run;
```

Figure 6: Weight function of IPCW and AHR.



By default the macro will use PROC PHREG's default method (that of Breslow) to handle ties:

```
%PSHREG(data=follicties, time=dftimeties, cens=cens, varlist=agedecade hgb clinstg chemo);
```

It is also possible to use the method of Efron, specifying `options=%str(ties=efron)`. In this example, the results differ slightly between these two ties handling methods:
Breslow method:

Analysis of Maximum Likelihood Estimates

Parameter	DF	Parameter Estimate	Standard Error	StdErr Ratio	Chi-Square	Pr > ChiSq	Hazard Ratio
agedecade	1	0.12793	0.04766	0.952	7.2044	0.0073	1.136
hgb	1	0.00688	0.00428	0.955	2.5814	0.1081	1.007
clinstg	1	0.40627	0.14321	0.967	8.0476	0.0046	1.501
chemo	1	-0.58711	0.19526	0.973	9.0405	0.0026	0.556

Efron method:

Analysis of Maximum Likelihood Estimates

Parameter	DF	Parameter Estimate	Standard Error	StdErr Ratio	Chi-Square	Pr > ChiSq	Hazard Ratio
agedecade	1	0.13443	0.04974	0.993	7.3058	0.0069	1.144
hgb	1	0.00730	0.00451	1.004	2.6128	0.1060	1.007
clinstg	1	0.43528	0.15159	1.020	8.2454	0.0041	1.545
chemo	1	-0.61705	0.20182	1.005	9.3476	0.0022	0.540

4.6 Stratification

If stratification is desired, two steps are necessary. First, it may be useful to stratify the Fine-Gray weights by the stratification variable, using the `cengroup` option. Assume that we would like to stratify the analysis by clinical stage (`clinstg`). We first specify

```
proc sort data=follic;
    by clinstg;
run;

%PSHREG(action=code, data=follic, time=dftime, cens=cens, varlist=agedecade hgb chemo,
        cengroup=clinstg);
```

To estimate the stratified model, we copy the `PROC PHREG` code from the Log window into the Editor and define the stratification variable `clinstg` in a `strata` statement (output not shown):

```
proc phreg covs(aggregate) data=dat_crr;
    model (_start_,_stop_)*_censcrr_(0)=agedecade hgb chemo;
    id _id_;
    weight _weight_;
    strata clinstg;
run;
```

4.7 Model-based estimation of cumulative incidence functions

In the following we illustrate how predicted cumulative incidence functions at different ages can be plotted, holding all other variables fixed at their means. Here the cumulative incidence function for the 25^{th}, 50^{th} and 75^{th} percentiles of age should be drawn, while the haemoglobin value, the clinical stage and the treatment with chemotherapy are fixed.

```
proc means data=follic;
    var age hgb clinstg chemo;
    output out=follicmeans mean=age hgb clinstg chemo;
```

```
run;

proc means data=follic NOPRINT;
    var age;
    output out=percentiles P25=perc25 P50=med P75=perc75;
run;

data percentiles;
    set percentiles;
    call symput("p25", perc25);
    call symput("median", med);
    call symput("p75", perc75);
run;

data follicmeans;
    set follicmeans;
    age=&p25; output;
    age=&median; output;
    age=&p75; output;
run;
```

After the percentiles have been computed and saved we run `%PSHREG(data=follic, time=dftime, cens=cens, varlist=age hgb clinstg chemo, action=code);`, copy the generated code from the Log to the Editor window, and modify the code as follows:

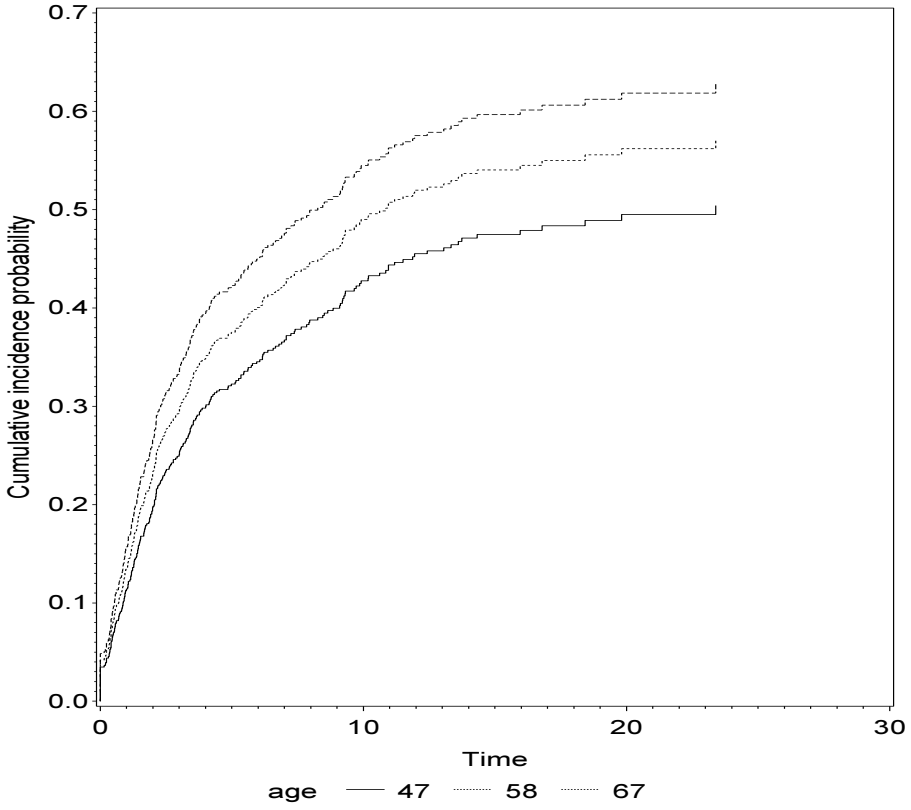
```
proc phreg covs(aggregate) data=dat_crr ;
    model (_start_,_stop_)*_censcrr_(0)=age hgb clinstg chemo;
    weight _weight_;
    baseline out=cuminccurves covariates=follicmeans survival=_surv_;
run;
```

To obtain a survival function estimate for each percentile it is necessary to input the before-computed percentiles of the variable `age` and the means of the remaining covariables as `covariates` in the `baseline` statement.
After running PROC PHREG in that way, we generate the cumulative incidence estimate by 1 minus the pseudo-survival function estimate. These can subsequently be plotted.

```
data cuminccurves;
    set cuminccurves;
    cuminc=1-_surv_;
run;

goptions reset=all;
```

```
symbol1 v=none i=steplj c=black line=1;
symbol2 v=none i=steplj c=black line=2;
symbol3 v=none i=steplj c=black line=3;
axis1 label=(angle=90 'Cumulative incidence probability');
axis2 label=('Time');

proc gplot data=cuminccurves;
    plot cuminc*_stop_=age /vaxis=axis1 haxis=axis2;
run;
```

Figure 7: Predicted cumulative incidence functions at ages 47, 58 and 67 adjusted by the means of haemoglobin value, clinical stage and treatment indicator.



4.8 Monotone Likelihood

To demonstrate the ability of our macro to deal with monotone likelihood, we artificially generate a subset of our data set, in which monotone likelihood occurs. In the first step a random variable `ranvar` is created and used to sort the data randomly. Then, we keep only the first 120 observations to reduce sample size.

```
proc sort data=follic; by stnum; run;

data follicmono;
```

```
    set follic;
    ranvar=ranuni(67);
run;

proc sort data=follicmono; by ranvar; run;

data sample; set follicmono (obs=120); run;
```

For the subset we now generate artificial event indicators, using percentiles of the random variable generated above, in such a way that we arrive at a data set where a standard Fine-Gray analysis would end up in monotone likelihood.

```
proc means data=sample NOPRINT;
    var ranvar;
    where chemo=1;
    output out=summary1 P50=med;
run;

data summary1;
    set summary1;
    call symput("median1", med);
run;

proc means data=sample NOPRINT;
    var ranvar;
    where chemo=0;
    output out=summary2 P25=perc25 P50=med P75=perc75;
run;

data summary2;
    set summary2;
    call symput("p25", perc25);
    call symput("median2", med);
    call symput("p75", perc75);
run;

data sample;
    set sample;
    if chemo=1 & ranvar < &median1 then crrmono=0;
    if chemo=1 & ranvar >= &median1 then crrmono=2;
    if chemo=0 & ranvar <= &median2 then crrmono=0;
    if chemo=0 & ranvar >= &median2 & ranvar < &p75 then crrmono=2;
```

```
    if chemo=0 & ranvar >= &p75 then crrmono=1;
run;

proc freq data=sample; table chemo crrmono chemo*crrmono; run;
```

True, when we use the `%PSHREG` macro with default settings the phenomenon of monotone likelihood occurs, implying non-convergence of the parameter estimates. To handle this it is reasonable to apply the Firth correction. This can be done with the macro parameter `FIRTH=1`. To obtain confidence limits of the hazard ratios the `rl` option has to be included in the code. As recommended by Heinze & Schemper (2001), we estimate profile penalized likelihood confidence limits by the additional option `options=%str(rl=pl)`, which will be directly passed to PROC PHREG's model statement:

```
%PSHREG(data=sample, time=dftime, cens=crrmono, varlist=agedecade hgb clinstg chemo,
        options=%str(rl=pl), Firth=1);
```

Below we can see the results from maximum likelihood analysis (no Firth correction) and from Firth-corrected analysis. The maximum likelihood parameter estimate of `chemo` is minus infinite, such that its hazard ratio is 0. The Firth correction provides an efficient way to deal with this problem; it arrives at a hazard ratio estimate of roughly 0.1.

Maximum likelihood analysis:

Parameter	HR	95% Confidence Limits			
		Robust		Wald	
		lower	upper	lower	upper
agedecade	1.305	1.024	1.664	0.975	1.747
hgb	0.994	0.969	1.019	0.968	1.021
clinstg	0.626	0.196	2.000	0.219	1.789
chemo	0.000	0.000	0.000	0.000	.

Firth-corrected analysis:

Parameter	HR	95% Confidence Limits			
		Wald		Profile Likelihood	
		lower	upper	lower	upper
agedecade	1.302	0.973	1.744	0.982	1.756
hgb	0.994	0.968	1.020	0.968	1.020
clinstg	0.686	0.247	1.905	0.227	1.761
chemo	0.099	0.006	1.747	0.001	0.708

The latter table does not yet provide a p-value for testing the hypothesis that chemo has no effect on the subdistribution hazard. Such a test can be obtained by an approximate penalized likelihood ratio. In penalized estimation, the penalized likelihoods of two nested models can not directly be compared. However, approximately the penalized likelihood ratios of two such models can be compared, because in each of the penalized likelihood ratios, the null likelihood is adequately penalized. (Because of the fact that the standard error ratios, relating robust and model-based standard errors, are about 1, it is reasonable to compare likelihoods; see also Geskus, 2011). For the model including all four variables, the global test statistics (on four degrees of freedom) are:

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	10.7171	4	0.0299
Score	8.9029	4	0.0636
Wald	6.8916	4	0.1417

Excluding chemo, we get:

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	4.6800	3	0.1968

```
Score                    4.5458        3          0.2082
Wald                     4.3786        3          0.2234
```

Under the null hypothesis that `chemo` has no effect on the cumulative incidence of relapse, the difference in likelihood ratio statistics is approximately χ^2 distributed with one degree of freedom. The difference of the test statistics and its p-value can be computed by

```
data plrtest;
    plrdiff=(10.7171-4.6800);
    pval=1-probchi(plrdiff,1);
    output;
run;
proc print;
run;
```

We obtain a significant p-value (0.014) which is in line with the profile (penalized) likelihood confidence interval for the effect of `chemo`:

```
Obs    plrdiff      pval

 1      6.0371    0.014008
```

5 Comparison with the R package cmprsk

For fitting proportional subdistribution hazards model, our SAS macro offers the same functionality as the `crr` function of the R package `cmprsk` (Gray, 2011). In addition, `%PSHREG` is also able to compute scaled Schoenfeld-type residuals, to apply the Firth correction, to compute profile likelihood confidence intervals, and to apply weighted estimation in case of time-dependent effects.

We also incorporated an option to specify how tied times to competing events and censoring times should be handled. Usually, one would assume that censoring occurs shortly after an event; this assumption can be consistently incorporated by the `tiedcens=after` option, which is the default in `%PSHREG`.

6 Comparison with the EVENTCODE option in SAS/STAT 13.1

Very recently, a new SAS version 9.4 (including SAS/STAT version 13.1) has been released in which the Fine-Gray model has been made directly available in PROC PHREG, by specifying the code of the event of interest in a new option EVENTCODE of the MODEL statement. All other codes which are not contained in the list of censoring values are then treated as competing event codes. We have compared the functionality of this new option with our macro by re-analyzing our examples. Even with the new EVENTCODE option, it is not possible to:

- predict cumulative incidence, neither for the whole sample nor at specific covariate values,
- apply variable selection (e.g., backward elimination),
- compute Schoenfeld-type residuals,

- apply the Firth correction or compute profile-likelihood based confidence intervals,
- use the ASSESS statement for assessing model assumptions using martingale residuals,
- include frailty effects.

All these options are possible with `%PSHREG` as it first modifies the input data set which can then be treated as any other survival data set, making full use of the functionality of PROC PHREG.

7 Availability, license and disclaimer

The macro is available under a GNU GPL license, version 2, at
http://cemsiis.meduniwien.ac.at/en/kb/science-research/software/statistical-software/PSHREG.

The license text can be accessed at http://www.gnu.org/licenses/gpl-2.0.txt.

References

Andersen, P. K., Geskus, R. B., de Witte, T., & Putter, H. (2012). Competing risks in epidemiology: possibilities and pitfalls. *International Journal of Epidemiology* **41**, 861–870.

Bakoyannis, G. & Touloumi, G. (2012). Practical methods for competing risk data: A review. *Statistical Methods in Medical Research* **21**, 257–272.

Durrleman, S. & Simon, R. (1989). Flexible regression models with cubic splines. *Statistics in Medicine* **8**, 551–561.

Fine, J. P. & Gray, R. J. (1999). A proportional hazards model for the subdistribution of a competing risk. *Journal of the American Statistical Association* **94**, 496–509.

Firth, D. (1993). Bias reduction of maximum likelihood estimates. *Biometrika* **80**, 27–38.

Geskus, R. B. (2011). Cause-specific cumulative incidence estimation and the Fine and Gray model under both left truncation and right censoring. *Biometrics* **67**, 39–49.

Grambauer, N., Schumacher, M., & Beyersmann, J. (2010). Proportional subdistribution hazards modeling offers a summary analysis, even if misspecified. *Statistics in Medicine* **29**, 875–884.

Gray, B. (2011). cmprsk: Subdistribution analysis of competing risks. *http://CRAN.R-project.org/package=cmprsk. R package version 2.2-2* .

Heinze, G. & Schemper, M. (2001). A solution to the problem of monotone likelihood in Cox regression. *Biometrics* **57**, 114–119.

Kalbfleisch, J. & Prentice, R. (1980). *The statistical analysis of failure time data.* John Wiley & Sons, New York.

Lin, D. & Wei, L. (1989). The robust inference for the Cox proportional hazards model. *Journal of the American Statistical Association* **84**, 1074–1078.

Pintilie, M. (2006). *Competing Risks - A Practical Perspective.* John Wiley & Sons, Chichester.

Pintilie, M. (2007). Analysing and interpreting competing risk data. *Statistics in Medicine* **26**, 1360–1367.

Rosthøj, S., Andersen, P., & Abildstrom, S. (2004). SAS macros for estimation of the cumulative incidence functions based on a Cox regression model for competing risks survival data. *Computer Methods and Programs in Biomedicine* **74**, 69–75.

Royston, P. & Altman, D. G. (1994). Regression using fractional polynomials of continuous covariates: Parsimonious parametric modelling. *Journal of Applied Statistics* **43**, 429–467.

SAS Institute INC (2010). *SAS/STAT Software.* Version 9.3. Cary, NC.

Schemper, M., Wakounig, S., & Heinze, G. (2009). The estimation of average hazard ratios by weighted Cox regression. *Statistics in Medicine* **28**, 2473–2489.

Therneau, T. M. & Grambsch, P. M. (2000). *Modeling Survival Data: Extending the Cox Model.* Springer, New York.

Xu, R. & O'Quigley, J. (2000). Estimating average regression effect under non-proportional hazards. *Biostatistics* **1**, 423–439.