



Weighted Cox Regression Using the R Package `coxphw`

Daniela Dunkler
Medical University
of Vienna

Meinhard Ploner
data-ploner.com

Michael Schemper
Medical University
of Vienna

Georg Heinze
Medical University
of Vienna

Abstract

Cox's regression model for the analysis of survival data relies on the proportional hazards assumption. However, this assumption is often violated in practice and as a consequence the average relative risk may be under- or overestimated. Weighted estimation of Cox regression is a parsimonious alternative which supplies well interpretable average effects also in case of non-proportional hazards.

We provide the R package `coxphw` implementing weighted Cox regression. By means of two biomedical examples appropriate analyses in the presence of non-proportional hazards are exemplified and advantages of weighted Cox regression are discussed. Moreover, using package `coxphw`, time-dependent effects can be conveniently estimated by including interactions of covariates with arbitrary functions of time.

Keywords: effect size, proportional hazards model, R package, survival analysis, weighted estimation.

1. Introduction

When analyzing survival data Cox's proportional hazards regression (Cox 1972) continues to be one of the most popular methods of analysis. However, its assumption of proportional hazards, i.e., the assumption that the effects of all covariates do not change over time, is often violated in practice. Consequently, the relative risk for a covariate exhibiting non-proportional hazards obtained by Cox regression may be under- or overestimated. But various methods of analysis are available to avoid this bias. Choosing an appropriate method may depend on

1. the aim of the analysis, i.e., whether a detailed description of the time-dependent effect or a single effect size measure is required,

2. the availability of a-priori subject matter information concerning a possible time-dependent effect,
3. the number of events and the sample size as it may restrict the complexity of the fitted model,
4. whether the covariate with non-proportional hazards is of primary interest or if it is used as an adjustment variable.

In the following, we briefly discuss methods, which can be used irrespective of the type of departure from proportional hazards, and for which implementations in R (R Core Team 2017) are readily available.

Often, *non-proportional hazards* are just *ignored* and standard Cox regression is used to obtain a single estimate of the hazard ratio (HR) for each covariate. This is then interpreted as an average effect of that covariate. During some periods of the follow-up this value will overestimate the true time-dependent effect and for other periods it will underestimate it. Such an analysis may be acceptable, if for a covariate a time-dependent effect has been detected with HRs constantly larger or smaller than 1, i.e., if the sign of the log HR estimate $\hat{\beta}(t)$ remains unchanged throughout follow-up. However, estimates from Cox regression ignoring non-proportionality were shown to be sensitive not only to the type of departure from proportionality but also to the censoring pattern of the data (Xu and O’Quigley 2000; Dunkler, Schemper, and Heinze 2010). Therefore, such an estimate does not generalize to its assumed uncensored population value.

The non-proportionality of one or more covariates can be taken into account explicitly by using either stratification, estimating an “extended Cox regression” where some coefficients are allowed to vary with time, or by using weighted Cox regression.

If only one categorical covariate with a small number of distinct categories which is not of primary interest exhibits non-proportional hazards then *stratification* might be suitable. This has the advantage that the type of time-dependent effect does not have to be specified. However, the effect of the stratifying covariate cannot be examined.

If one assumes that a covariate of primary interest has different short-term and long-term effects, one could express these effects by *estimating piecewise constant HRs* for consecutive periods using separate models. However, such an analysis is based on the assumptions of constant HRs within each period and a sudden change at the cutpoint between two periods. Both assumptions are usually unrealistic. But such a piecewise constant model might sometimes be preferred over more complex approaches because of its simplicity.

More generally, Cox regression can be extended to accommodate different types of non-proportional hazards by *including an interaction of a covariate with time*. This is technically realized by a time-dependent covariate, which is a product of the value of a covariate with a pre-specified function of time t , $\gamma(t)$. The function $\gamma(t)$ can be chosen such that a time-dependent effect of a particular shape results. With $\gamma(t) = t$, a linear time-dependent effect will be obtained, while $\gamma(t) = \log(t)$ will yield a log-linear time-dependent effect. More flexible time-dependent effects can be obtained, if products of the covariate with the elements of a data-derived vector-valued function of time are used (Lehr and Schemper 2007; Royston and Altman 1994; Heinzl and Kaider 1997). Significant regression coefficients for these interaction terms suggest that the respective covariate does not exhibit proportional hazards. A multiple Cox regression extended in this way may even include several interactions of various

covariates with time. This option of analysis is the most flexible one, however, it is only useful with larger sample sizes and if a concise description of the time-dependent effect is of interest. Furthermore, it will not always be possible to draw clear decisions from such models, in particular, if the sign of the corresponding regression coefficient changes during follow-up. For example, in a medical context, the purpose of a model with survival data is often to decide which therapy most likely will lead to longer survival. Using an extended Cox regression model with a time-by-covariate interaction to account for non-proportional hazards may not be suitable to answer this question.

Finally, *weighted Cox regression* is a parsimonious option to account for time-dependent effects applicable also for small samples sizes (Schemper 1992; Sasieni 1993). Schemper, Wakounig, and Heinze (2009) suggested to weight the contributions of each event time to the partial likelihood according to the product of the survivor function and the inverse cumulative probability of follow-up at that time. Therefore, at each event time, their weighting function is proportional to the expected number of subjects at risk if censoring had not occurred. By applying these weights, weighted Cox regression yields estimates of *average HRs* which themselves approximate the *odds of concordance*, defined for two treatment groups A and B as $OC = P(T_A < T_B)/P(T_B < T_A)$, where T_A and T_B denote survival times from two subjects randomly selected from these groups. Odds of concordance can be conveniently transformed into concordance probabilities $c = P(T_A < T_B) = OC/(OC + 1)$, which are intuitive effect size measures. They permit a clear-cut answer to the question which treatment or level of a prognostic factor is preferable, independent from the assumption of proportional hazards.

Here we present the R package **coxphw** (Heinze, Ploner, and Dunkler 2018) which implements weighted estimation in Cox regression providing unbiased average HR estimates irrespective of proportionality of hazards. Moreover, our package provides options to estimate time-dependent effects conveniently by including interactions of covariates with arbitrary functions of time, with or without making use of the weighting option. Two biomedical examples are introduced in Section 2 motivating and exemplifying the application of weighted Cox regression. A brief review of the weighted Cox regression methodology is provided in Section 3, followed by the description of the R package **coxphw** in Section 4. A small simulation study illustrates the beneficial effect of weighted Cox regression (Section 5) before in Section 6 weighted and standard Cox regression are applied to the examples and differences in estimates and interpretation are discussed. Final remarks follow in Section 7.

2. Motivating examples

2.1. Gastric cancer study

The **Gastrointestinal Tumor Study Group** (1982) presented results from a clinical trial on survival of 90 patients with locally advanced, non-resectable gastric carcinoma comparing treatment by a combination of chemotherapy and radiation to chemotherapy alone (variable **radiation**). The aim of this randomized study was to determine if radiation is beneficial to the survival of patients with gastric carcinoma under chemotherapy. During a median follow-up time of 4.34 years 79 (87.78%) events were recorded; in the groups with and without radiation 42 (93.33%) and 37 (82.22%) patients died, respectively. The data set is available in the R package **coxphw**.

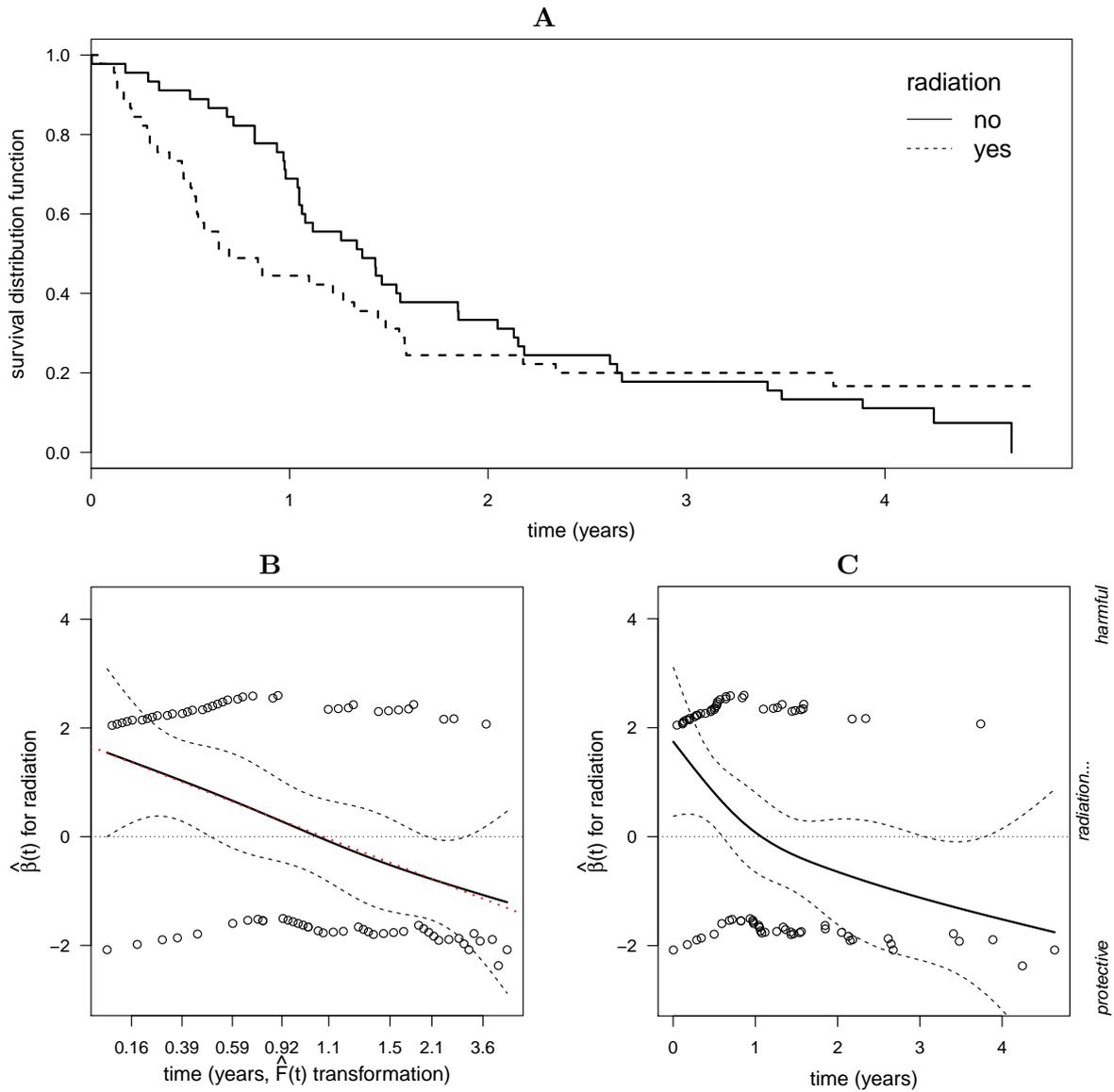


Figure 1: Gastric cancer study: (A) Plot of cumulative survival probabilities for **radiation**. Crossing survival curves indicate non-proportional hazards, i.e., a time-dependent effect. (B) A plot of scaled Schoenfeld residuals versus $\hat{F}(t)$ as a means to detect non-proportionality. The circles represent the scaled Schoenfeld residuals. The black continuous line is the locally weighted scatter-plot smoother (LOWESS) of the scaled Schoenfeld residuals and the black dashed lines are its 95% confidence limits. The red dotted line is a linear regression line of the scaled Schoenfeld residuals on the transformed time. (C) To directly visualize the time-dependent effect $\hat{\beta}(t)$ of **radiation** a plot of scaled Schoenfeld residuals versus time is usually preferred. However, during time periods with few events the log relative hazard $\hat{\beta}(t)$ should not be overinterpreted.

Crossing cumulative survival curves as visualized in Figure 1A are typical for situations of non-proportional hazards. In this example, one may hypothesize that radiation has a harmful effect, but only within the first two years. Generally, plots of scaled Schoenfeld residuals versus (functions of) time t can be used to detect non-proportionality (Grambsch and Therneau 1994) (Figure 1B). Following Grambsch and Therneau (1994), the locally weighted scatter-plot smoother (LOWESS) applied to the scaled Schoenfeld residuals approximates the time-dependent log relative hazard $\hat{\beta}(t)$ of treatment with radiation (`radiation = 1`) versus treatment without radiation (`radiation = 0`, reference). In this plot, we have used the Kaplan-Meier estimate of the empirical distribution function $\hat{F}(t)$ for the x -axis instead of time t itself, which has the advantage that observed events are evenly distributed over the axis. Consequently, the LOWESS is not disproportionately influenced by periods during which the occurrence of events is rare. In case of proportional hazards, one would expect the LOWESS to follow a horizontal line. Therefore, in Figure 1B any departure from a horizontal line indicates a violation of the proportional hazards assumption of `radiation`. Such a departure can be assessed by testing, in a linear regression analysis of scaled Schoenfeld residuals on time, whether the slope estimate is different from zero (Harrell 2001, p. 487). For `radiation` this test, which is also implemented in the `cox.zph` function of the `survival` package (Therneau 2017), gives a highly significant p value < 0.001 . Generally, such a test may fail to reveal departures from proportionality in case of non-monotone, e.g., U-shaped, trends. From Figure 1C, a plot of scaled Schoenfeld residuals versus untransformed times, we learn that the effect of radiation may change slightly faster during earlier times.

Ignoring the time-dependent effect and applying a Cox regression (using the `coxph` function of the R package `survival`) gives a HR for treatment with radiation of 1.15 (95% CI: 0.74–1.81; $p = 0.537$) compared to treatment without radiation. However, this effect estimate is biased if the true HR is time-dependent, and likely also biased with respect to the true average HR. The amount of this latter bias depends on the combination of the time-dependency and the censoring pattern in this study. In Section 6.1 we will revisit this example and discuss other, more appropriate methods of analysis which will take the time-dependent effect of `radiation` correctly into account.

2.2. Biofeedback therapy study

In this small clinical trial comprising 33 patients suffering from aspiration after head and neck surgery the effect of biofeedback therapy (`bfb`) on time until success of therapy, i.e., complete swallowing rehabilitation, was evaluated (Denk and Kaider 1997). Patients were randomized into two groups: 19 (57.58%) patients received videoendoscopic biofeedback therapy, i.e., the patients and therapists could visually control swallowing maneuvers on a monitor (`bfb = 1`) and 14 (42.42%) received the standard therapy (`bfb = 0`). The median therapy duration was 25 days; during follow-up 13 (68.42%) events of successful swallowing rehabilitation were recorded in the biofeedback group and 10 (71.43%) events were observed in the group without biofeedback. The therapy was started as soon as the healing process after surgery was finished. The time elapsing from surgery to start of therapy was considered an important covariate determining success of therapy. Because of the skewed distribution of this time interval, a \log_2 -transformation was applied prior to analysis (variable `log2heal`). The data set is available in the R package `coxphw`.

Kaplan-Meier analysis provides some graphical evidence that an apparent early advantage of

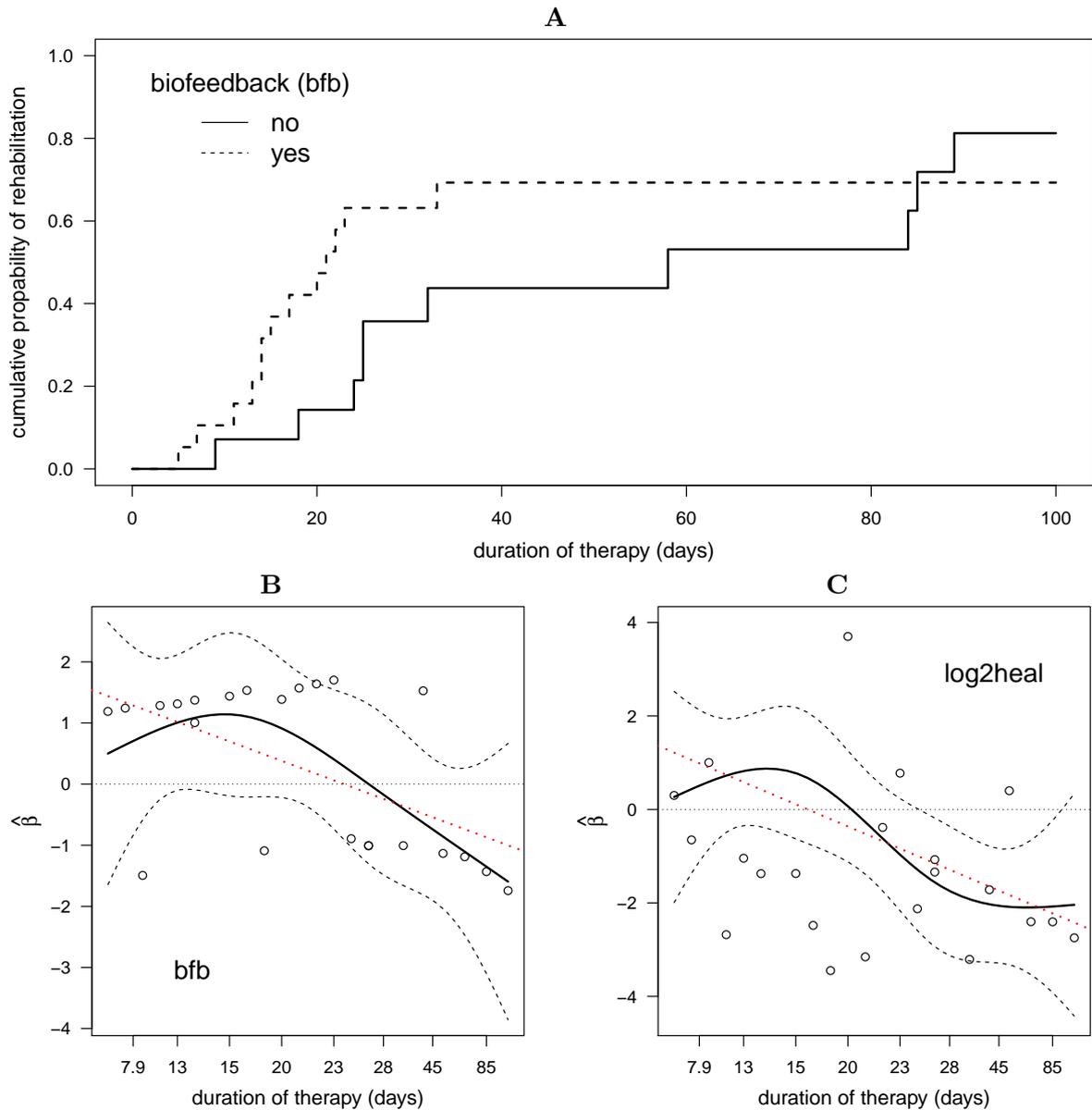


Figure 2: Biofeedback therapy study: (A) Cumulative survival plot depicting the probability of full swallowing rehabilitation due to `bfb`. Since `log2heal` is not accounted for in the Kaplan-Meier analysis, the plot should not be the only means to check the proportional hazards assumption of `bfb`. Based on a multiple Cox regression with `bfb` and `log2heal`, plots of scaled Schoenfeld residuals versus $\hat{F}(t)$ for (B) `bfb` and (C) `log2heal` are shown, as a means to detect non-proportionality. For more information on the plots see the caption of Figure 1B. The range of the y -axis of the plot with `log2heal` is restricted to -4 to 4 .

biofeedback therapy may vanish with ongoing duration of therapy (Figure 2A). Hence, the assumption of proportional hazards could be violated for `bfb`. The plots of scaled Schoenfeld residuals from a multiple Cox regression with `bfb` and `log2heal` versus $\hat{F}(t)$ visualize possible time-dependent effects of both covariates (Figures 2B and 2C). (Non-proportional hazards

should always be checked in the multiple Cox regression model as the confounding of a covariate with other covariates can induce or remove non-proportional hazards.) A global test for non-proportional hazards of any covariate, as implemented in `cox.zph` of the `survival` package, gives a p value of 0.001, and for `bfb` and `log2heal` the p values are 0.036 and 0.003, respectively. Figures 2B and 2C only approximate the shapes of the time-dependent effects, because in each case there was no adjustment for a time-dependent effect of the other covariate in the model.

Time-dependent effects of `bfb` and `log2heal` could be modeled by an extended Cox regression including additional time-dependent covariates, i.e., products of `bfb` and `log2heal` with suitable functions of time. However, small sample size and small number of events restrict the number of degrees of freedom which can be used during model development considerably and hence impose restrictions on the complexity of the fitted model (Steyerberg, Eijkemans, Harrell, and Habbema 2000).

In Section 6.2 we shall discuss how an average log hazard ratio of `bfb` can be estimated and how the estimate can be transformed into a simple effect size measure for clear-cut therapy decisions.

3. Methods

3.1. Weighted Cox regression

In a sample of N ($1 \leq i \leq N$) subjects m distinct and uncensored survival times t_j ($m \leq N$, $1 \leq j \leq m$) are observed. For each subject i we assume that a covariate (row) vector x_i of k ($r = 1, \dots, k$) covariates is recorded. The set of subjects without the event and uncensored prior to t_j , i.e., the risk set, is denoted by R_j . Using a Cox regression a vector of coefficients is estimated by setting the first derivative of the log partial likelihood (the score vector) to zero, and by solving the resulting estimating equations numerically, usually using the iterative Newton-Raphson algorithm. In weighted Cox regression a weight function $w(t_j)$ is introduced to weight the contributions to the partial likelihood at the m uncensored failure times differently. Thus, weighted regression coefficients $\hat{\beta}$ are derived by solving the estimating equations

$$\frac{\partial \log L(\beta)}{\partial \beta_r} = \sum_{j=1}^m w(t_j) \left[x_{jr} - \frac{\sum_{l \in R_j} x_{lr} \exp(\hat{\beta}^\top x_l)}{\sum_{l \in R_j} \exp(\hat{\beta}^\top x_l)} \right] = 0.$$

If the weights $w(t_j)$ are set to 1, then a standard Cox regression is obtained. Early versions of weighted Cox regression were proposed by Schemper (1992) and Sasieni (1993). These suggestions used as weights either the observed number of individuals at risk $R(t_j)$ or the empirical survivor function estimate at these times $\hat{S}(t_j)$, extending the tests by Breslow (1970) and Prentice (1978), respectively, to a multi-covariate situation as does the Cox model to Mantel's logrank test (Mantel 1966). In these suggestions, weighted Cox regression was proposed as a robust alternative to the standard Cox estimator, reducing the influence of a violated proportional hazards assumption, but also of outlying survival times on parameter estimates. Xu and O'Quigley (2000) showed that without censoring standard Cox regression estimates a time-averaged effect also in case of non-proportional hazards. Let $G(t_j)$ denote

the cumulative probability of follow-up until t_j , i.e., the survivor function of the potential follow-up time distribution. Then, applying the weighting function $w(t_j) = G(t_j)^{-1}$ to the score equations yields time-averaged regression effects also for censored samples (Xu and O’Quigley 2000).

Finally, Schemper *et al.* (2009) suggested the weighting function $w(t_j) = S(t_j)G(t_j)^{-1}$. These weights are proportional to $N_E(t_j)$, the expected number of subjects at risk at each t_j if there had been no censoring, and this greatly enhances interpretability of the resulting hazard ratio estimates, denoted as “average hazard ratios” by Schemper *et al.* (2009). Specifically, these authors have shown that in a two-group comparison, average hazard ratios approximate the odds of concordance $OC = c/(1 - c) = P(T_A < T_B)/P(T_B < T_A)$ very well, T_A and T_B being the survival times of two randomly chosen subjects of groups A and B . Intuitively, this can be explained as *each pairwise comparison* of a subject that fails with all subjects still at risk being given equal importance. (Approximately, if a risk set has N_E subjects and one of them fails, there are $N_E - 1 \approx N_E$ pairwise comparisons within the risk set.) By contrast, standard Cox regression attributes similar importance to *each risk set*. Further details are contained in the contribution by Schemper *et al.* (2009).

The approximation extends to the general situation of a continuous covariate x , where a *generalized concordance probability* c' can be defined as

$$c' = P(T_i < T_e | x_i = x_e + 1),$$

where T_i and T_e the survival times of randomly chosen subjects with covariate values x_i and x_e , respectively. Generalized odds of concordance follow directly as $c'/(1 - c')$. The definition of c' also includes the *concordance probability* c of a two-group comparison, setting $x_i = 1$ and $x_e = 0$. The approximation of the odds of concordance with average HRs is valid independent of the type of non-proportionality and the censoring pattern. Thus, c' is useful if the analyst is not interested in a detailed description of a time-dependent effect, but in a summary measure suitable for decision making. If proportional hazards can be safely assumed, an estimate of c' (and of c) can also be derived from standard Cox regression by (Dunkler *et al.* 2010)

$$\hat{c}' = \exp(\hat{\beta}) / (1 + \exp(\hat{\beta})).$$

Weighted Cox regression models can be estimated by plugging estimates for $S(t_j)$ and $G(t_j)$ into the weighting function, such as $\hat{S}(t_j)$, the left-continuous Kaplan-Meier estimate of the survivor function at time t_j , and $\hat{G}(t_j)$, the probability of still being followed-up at t_j , estimated by the Kaplan-Meier method with reversed meaning of the status indicator (Schemper and Smith 1996). Inference for estimates of weighted Cox regression can be based on the Lin (1991) and Sasieni (1993) sandwich estimate $\hat{A}^{-1}\hat{B}\hat{A}^{-1}$ with $-\hat{A}$ and $-\hat{B}$ denoting the sum of contributions to the second derivative of the log likelihood, weighted by $\hat{w}(t_j)$ and $\hat{w}(t_j)^2$, respectively. This estimate is independent from the scaling of the weights and in case of no weighting it is algebraically equal to the usual variance estimate based on inversion of the Fisher information matrix. This variance estimate is theoretically valid only in case of proportional hazards and without model misspecification. Therefore, and since application of weighted Cox regression usually implies a violated proportional hazards assumption, the robust covariance estimate should be used (Lin and Wei 1989; Therneau and Grambsch 2000, Chapter 7.2). An estimate of the covariance matrix can also be obtained by using the Jackknife, where the regression coefficients are computed leaving out each individual in turn.

The resulting matrix of “difference in $\hat{\beta}$ ” residuals \hat{D} is then used to compute the covariance matrix as $\hat{V} = \hat{D}^\top \hat{D}$. The Jackknife variance estimate is valid also in case of misspecification. Unequal weighting of the contributions to the likelihood usually increases the variability of estimates, in particular in situations of proportional hazards. However, empirical studies have shown that this increase appears to be moderate (Wakounig, Heinze, and Schemper 2015), and this finding is further supported by our simulation study presented in Section 5.

4. The R package `coxphw`

The R package `coxphw` implements weighted estimation of Cox regression. The most important arguments of the package’s main function `coxphw` are:

- `formula` is an R formula with a survival object as returned by the `Surv` function from the `survival` package (Therneau 2017) as the response on the left of the `~` operator, and the model terms on the right. In order to specify time-dependent effects, the model terms can also contain products of covariates with functions of the survival time variable (e.g., `Surv(t, status) ~ x + log(t):x`), which the formula parser of `coxphw` will identify and treat appropriately.
- `data` is a data-frame containing the variables given in `formula`.
- `template` determines which type of (weighted) estimation of Cox regression is requested. The analyst can choose between "AHR" for estimation of average HRs (Schemper *et al.* 2009), "ARE" for estimation of average regression effects (Xu and O’Quigley 2000) and "PH" for unweighted Cox proportional hazards regression. "AHR" is the default type of estimation.
- `robust` requests the robust (Lin-Wei) covariance matrix used for any significance testing, if set to `TRUE`. This is the recommended choice and the default.
- `jack` requests that the covariance matrix is estimated by the Jackknife. Default is set to `FALSE`.
- `trunc.weights` can be used to truncate the weights such that disproportional influence of some outlying event times is avoided. Weight truncation can increase the precision of estimates, but introduces a mild bias. `trunc.weights` specifies the quantile of the observed weight distribution at which weights are to be truncated. Before using this option, it is recommended to inspect the values of the weights as a function of time, using the `plot` function (see below). For mild truncation, a value of 0.95 is recommended. Default is 1, indicating that no truncation is applied.
- `alpha` is the significance level of confidence limits (confidence level $1 - \alpha$) with 0.05 as default.
- `betafix` can be used to hold some of the regression coefficients constant at pre-defined values, in particular for use with time-dependent effects. A vector with one element for each regression coefficient is expected. While an `NA` element will lead to estimation of the corresponding effect, a numeric element will hold the related regression coefficient constant at this value and prevent its estimation. The default value is `betafix =`

`NULL`, requesting unrestricted estimation of all regression coefficients. In Section 6.2 an application of this option will be presented. `betafix` has a similar meaning and can be used alternatively to specifying an `offset` term in the model `formula`. However, while `offset` is restricted to time-fixed effects, `betafix` enables the user to hold time-varying regression coefficients constant.

The package's two workhorse functions have been implemented in Fortran 90. The first function, `WEIGHTEDCOX`, applies Newton-Raphson iteration to arrive at the regression coefficients (Thisted 1988, p. 164). The second function, `LIKE`, computes the weighted first and second derivatives of the log partial likelihood, which are the main ingredients of the Newton-Raphson iteration. This latter function uses efficient updating of these quantities over the risk sets. `coxphw` provides some possibilities to fine-tune the iterative estimation procedure, which are explained in the help pages. `coxphw` uses Breslow and Crowley (1974)'s tie-handling method. A call to `coxphw` generates an object of class `'coxphw'`, and corresponding `summary`, `print`, `coef`, `vcov` and `confint` methods are available. A `plot` method can be used to visualize the weighting function. Further utility functions contained in the `coxphw` package are:

- `predict`, which computes potentially non-linear or time-dependent effect estimates at specified values of a variable and has an accompanying `plot` method for visualizing time-dependent or non-linear hazard ratios,
- `concord`, which computes the generalized concordance probabilities \mathcal{C}' for each covariate. This function can also be applied to an object of class `'coxph'`.

5. Simulation

We conducted a small simulation study to demonstrate similarity of standard and weighted Cox regression under proportional hazards, and the advantages of weighted Cox regression in case of non-proportional hazards. For simplicity, we simulated studies with one binary explanatory variable x , $x_i \sim \text{Bernoulli}(0.5)$, $i = 1, \dots, N$. In the proportional hazards scenario (Figure 3PH), we simulated survival times following a Cox-Weibull distribution (Bender, Augustin, and Blettner 2005), by first generating uniformly distributed random numbers $u \sim U[0, 1]$ and then computing survival times as $t = [-\log(u)/0.11 \exp(\beta^\top x)]^{1/1.22}$ with $\beta = 0.2007$. In the non-proportional hazards scenario (Figure 3NPH), we assumed that survival times of subjects with $x_i = 0$ followed a Cox-exponential distribution, obtained by $t = -\log(u)/0.3565$, while the survival times of those with $x_i = 1$ followed a Cox-Gompertz distribution, obtained by $t = (1/1.6) \log[1 - 1.6 \log(u)/0.0228]$. This setting was chosen so that both scenarios have underlying concordance probabilities of 0.55, which were numerically determined by $P(T_1 < T_0) = \int f_1(t)S_0(t)dt$, with $f_1(t)$ denoting the density of survival times in group of subjects with $x = 1$ and $S_0(t)$ denoting the survivor function in the group with $x = 0$ (Dunkler *et al.* 2010). In each scenario we simulated 2000 data sets with $N = 1000$ observations each. To quantify the effect of censoring, we simulated both administrative and exponential censoring. Administrative censoring was implemented by censoring all survival times t at a pre-defined maximum follow-up time τ if $t > \tau$. We iteratively determined τ to obtain proportions of administrative censored times of 10% and 15%. For exponential censoring a Cox-exponentially distributed potential follow-up time was simulated as $y =$

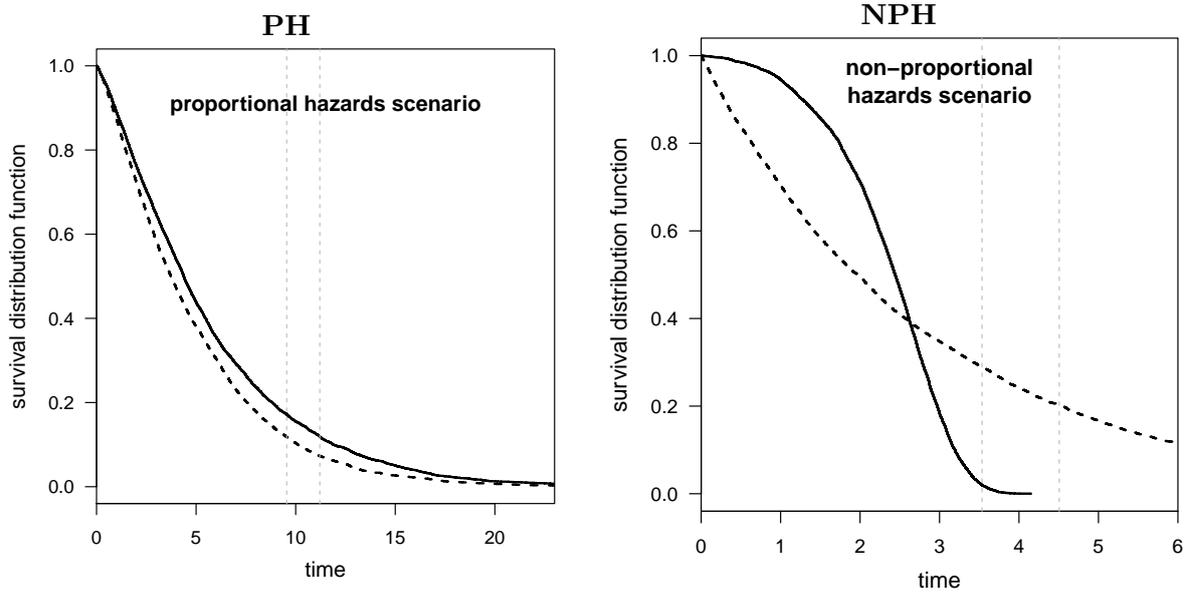


Figure 3: Simulation: Plot of cumulative survival probabilities of the proportional hazards scenario (PH) and the non-proportional hazards scenario (NPH). The gray vertical lines indicate the assumed maximum follow-up times for administrative censoring of 10% and 15%.

$-\log(u)/\lambda$, where $u \sim U[0,1]$ and $\lambda = 0.0602$ for the proportional hazards scenario and $\lambda = 0.122$ for the non-proportional hazards scenario, determining these values such that exponential censoring rates of 25% were obtained. The observable survival times were then defined as $t^* = \min(t, y)$. For each simulated data set we generated six versions; a data set with A) no censoring, B) 25% exponential censoring, C) 10% administrative censoring, D) both B and C, E) 15% administrative censoring, F) both B and E. With uncensored data the concordance probability $\hat{c} = P(T_i < T_e | x_i = 0, x_e = 1)$ is algebraically equal to the nonparametric two-sample test statistic of [Mann and Whitney \(1947\)](#) and to the area under the receiver operating characteristic curve ([Hanley and McNeil 1982](#)).

From [Table 1](#) we learn that in the proportional hazards scenario from both Cox regression and weighted Cox regression unbiased estimates of concordance probabilities can be derived, and Cox regression, which explicitly exploits the proportional hazards assumption, gives slightly more efficient estimates. However, in case of non-proportional hazards the Cox regression coefficients vary with the type and the magnitude of censoring, while weighted Cox regression coefficients remain fairly constant. When deriving estimates of concordance probabilities from the regression coefficients, we see that there is almost no systematic bias induced by weighted Cox regression, while Cox regression grossly underestimates c . While exponential censoring alone does not systematically impact the weighted Cox estimates, administrative censoring limits the time range covered by the data to a maximum value of τ and thus restricts the averaging of the time-dependent hazard ratio to the domain $t \in [0, \tau]$. In this case, no estimation procedure can yield unbiased estimates without making strong and untestable assumptions about the regression coefficients beyond τ .

Scenario (percent censored)	Cox regression		Weighted Cox regression	
	$\hat{\beta}$	$\hat{c} \times 100$	$\hat{\beta}$	$\hat{c} \times 100$
PH: A (0)	0.200	55.0 (0.025)	0.200	55.0 (0.028)
PH: B (25)	0.199	55.0 (0.029)	0.199	55.0 (0.030)
PH: C (10)	0.200	55.0 (0.026)	0.200	55.0 (0.029)
PH: D (29)	0.199	55.0 (0.029)	0.199	55.0 (0.031)
PH: E (15)	0.200	55.0 (0.027)	0.200	55.0 (0.029)
PH: F (32)	0.199	55.0 (0.030)	0.199	55.0 (0.031)
NPH: A (0)	-0.382	40.6 (0.028)	0.213	55.3 (0.033)
NPH: B (25)	-0.211	44.7 (0.031)	0.207	55.1 (0.036)
NPH: C (10)	-0.382	40.6 (0.028)	0.213	55.3 (0.033)
NPH: D (29)	-0.211	44.7 (0.031)	0.207	55.1 (0.036)
NPH: E (15)	-0.342	41.5 (0.029)	0.223	55.5 (0.033)
NPH: F (32)	-0.180	45.5 (0.032)	0.216	55.4 (0.035)

Table 1: Simulation results: Mean estimated regression coefficients $\hat{\beta}$ and concordance probabilities $\hat{c} \times 100$ by Cox regression and by weighted Cox regression for the proportional hazards scenario (PH) and the non-proportional hazards scenario (NPH) are shown. For each of the 2000 simulated data sets and for each scenario six versions were obtained and analyzed; A) no censoring, B) 25% exponential censoring, C) 10% administrative censoring, D) both B and C, E) 15% administrative censoring, F) both B and E. Under both PH and NPH scenarios and no censoring, the true concordance probability is 0.55. Standard errors for $\hat{\beta}$ were all 0.001 if rounded to the third decimal place. Standard errors for $\hat{c} \times 100$ are given in brackets.

6. Examples revisited

6.1. Gastric cancer study

In Section 2.1 we have seen that the effect of **radiation** on survival in the gastric cancer study suggests a time-dependent effect, i.e., an initially harmful effect of radiation decreasing over time. In the following, we describe the results according to various methods, focusing on the capabilities of package **coxphw**.

Ignoring non-proportional hazards

Ignoring non-proportionality a standard Cox regression can be estimated using **coxphw** by the following command:

```
R> coxphw(Surv(yrs, status) ~ radiation, data = gastric, template = "PH")
```

or, equivalently, using the R package **survival**:

```
R> coxph(Surv(yrs, status) ~ radiation + cluster(id), data = gastric,  
+ method = "breslow")
```

In **coxphw**, the robust variance estimate is computed by default to account for misspecifications such as lack of proportional hazards. In **coxph** the robust variance is invoked by

adding the term `cluster(id)` where `id` is a unique subject identifier. In `coxphw` by default each line is assumed a distinct “cluster” for the computation of the robust covariance, but an `id` argument is available to specify other clusters of observations if needed. Computation of the robust variance could also be entirely switched off by setting `robust = FALSE`. The estimated HR of `radiation`, obtained by both equivalent function calls, is 1.15 (95% CI: 0.74–1.81; $p = 0.537$). However, as can be seen in Figure 1B the sign of the time-dependent log HR $\hat{\beta}(t)$ of `radiation` changes during follow-up and consequently the non-proportional hazards of `radiation` should probably not be ignored. Since some survival times are censored, we must consider that the coefficient obtained by this standard Cox regression is possibly a biased estimate of the average population effect.

Estimating piecewise constant hazard ratios

As the cumulative survival probabilities (Figure 1A) and the Schoenfeld residuals (Figures 1B and 1C) suggest differing short-term and long-term effects of `radiation`, separate modeling of different time periods might be a more suitable method of analysis. In this example, we assume that no prior knowledge of the time-dependent effect is available which could guide the choice of an appropriate cutpoint. Thus, time periods are separated such that an equal number of events are observed before and after the cutpoint. In our example, this cutpoint is at one year. Two separate Cox regression models are now fitted, describing the effect of `radiation` during the first year, and after the first year. In the model for the first year any survival times longer than one year are censored at this point. Hence, the data set for this model consists of 90 patients experiencing 39 (43.33%) events during follow-up with 25 (55.56%) events in the radiation group and 14 (31.11%) events in the group without radiation. A plot of scaled Schoenfeld residuals (see accompanying R example code) still shows some non-proportional hazards but the test for non-proportionality now yields a p value of 0.103. For the second time period only subjects alive at one year and with a follow-up longer than one year are included, leading to a sample of 51 patients with 40 events (78.43%); with 12 (60.00%) and 28 (90.32%) events occurring in the groups with and without radiation, respectively. The test for non-proportionality gives a p value of 0.446.

The first Cox regression model reveals a significant harmful effect of radiation on survival (HR 2.40; 95% CI: 1.27–4.55; $p = 0.007$) up to one year. By contrast, the second model suggests a non-significant protective effect of radiation (HR 0.55; 95% CI: 0.28–1.08; $p = 0.081$) after one year.

Including a time-by-covariate interaction

Extending a Cox regression by an interaction of `radiation` with time, a single model could be used to estimate the effect of `radiation` in the two time periods. Using `coxphw` this model can be estimated very conveniently by first defining a function of time which represents the piecewise constant estimation, and secondly, by including a product term of `radiation` and this function of time in the `formula` of a call to `coxphw`. The function using a simple time-varying indicator to separate two periods after one year is defined by

```
R> fun <- function(t) as.numeric(t > 1)
```

The product of `radiation` and `fun(t)` is specified as a model term in the `formula` argument of `coxphw`:

```
R> coxphw(Surv(yrs, status) ~ radiation + fun(yrs):radiation,
+ data = gastric, template = "PH")
```

The `coxphw` function can recognize that the name of the time variable (`yrs`) appears in the product term. As a consequence, the function will compute `fun(yrs)` at each event time and generate a time-dependent product term, which for each subject is updated in each risk set. This way of specifying time-by-covariate interactions is even more convenient than that used by `coxph`, which uses a special argument, `tt`, to define such interactions:

```
R> coxph(Surv(yrs, status) ~ radiation + tt(radiation) + cluster(id),
+ tt = function(x, t, ...) x * (t > 1), data = gastric,
+ method = "breslow")
```

`coxphw` will yield the following estimates, which are identical to those obtained by the call to `coxph`:

Model fitted by unweighted estimation (PH template)

	coef	se(coef)	exp(coef)	lower0.95	upper0.95	z	p
radiation	0.877	0.326	2.405	1.270	4.554	2.693	0.007
fun(yrs):radiation	-1.483	0.476	0.227	0.089	0.577	-3.116	0.002

Wald Chi-square = 10.300 on 2 df, p = 0.006

The HR estimate (`exp(coef)`) of `radiation`, i.e., the effect of radiation compared to therapy without radiation, in the first period is 2.405, as in this period, `fun(yrs)` equals 0. The HR of `fun(yrs):radiation` has to be interpreted as the ratio of the two HRs of both time periods of `radiation`: the HR of the second period divided by the HR of the first period. Thus, the HR of the second period follows as the product of the two HRs, $2.405 \cdot 0.227 = 0.546$, or, equivalently, as the exponentiated sum of the two regression coefficients, $\exp(0.8774 - 1.4826) = \exp(-0.6052) = 0.546$.

An extended Cox regression may include interactions of covariates with any function of time $\gamma(t)$ and thus is able to accommodate time-dependent effects. The simplest function of time $\gamma(t)$ is $\gamma(t) = t$, which leads to the estimation of an effect linearly changing with time. It can be estimated with `coxphw` using the following function call:

```
R> fit1 <- coxphw(Surv(yrs, status) ~ radiation + yrs:radiation,
+ data = gastric, template = "PH")
```

or equivalently using `coxph`

```
R> coxph(Surv(yrs, status) ~ radiation + tt(radiation) + cluster(id),
+ tt = function(x, t, ...) x * t, data = gastric)
```

Again both function calls yield identical estimates.

Model fitted by unweighted estimation (PH template)

	coef	se(coef)	exp(coef)	lower0.95	upper0.95	z	p
radiation	1.270	0.433	3.561	1.522	8.328	2.930	0.003
yrs:radiation	-0.965	0.341	0.381	0.195	0.743	-2.831	0.005

Wald Chi-square = 9.048 on 2 df, p = 0.011

The regression coefficient of `radiation` (1.270) is the effect of chemotherapy at `yrs = 0`, i.e., at baseline. The coefficient corresponding to `yrs:radiation` (-0.965) gives the yearly change of this effect. From the two regression coefficients, the HRs at arbitrary time points can be derived by inserting different times into the formula: $\exp(1.270 - 0.965 \times \text{yrs})$. The HRs at 6 months, one year and two years are estimated as 2.20 (95% CI: 1.21–3.99), 1.36 (0.85–2.15), and 0.52 (0.24–1.12), respectively. These estimates can also be obtained conveniently by using the `predict` function of `coxphw`:

```
R> predict(fit1, type = "slice.time", x = "yrs", z = "radiation",
+   newx = c(0.5, 1, 2), exp = TRUE, pval = TRUE, verbose = TRUE)
```

	yrs	HR	HR lower 0.95	HR upper 0.95	p
1	0.5	2.198	1.210	3.992	0.0098
2	1.0	1.356	0.854	2.155	0.1971
3	2.0	0.517	0.238	1.121	0.0946

Objects obtained with `predict` can also be plotted to visualize the time-dependent effect (Figure 4).

```
R> fitlest <- predict(fit1, type = "slice.time", x = "yrs", z = "radiation",
+   newx = seq(from = 0.1, to = 3, by = 0.1))
R> plot(fitlest, addci = TRUE)
```

Often a log-linear function of time $\gamma(t) = \log(t)$ is more suitable to model time-dependent effects, in particular if the HR changes faster during early follow-up times compared to later times. Such a log-linear time-dependent effect can be requested by the following command:

```
R> coxphw(Surv(yrs, status) ~ radiation + log(yrs):radiation,
+   data = gastric, template = "PH")
```

In our example, this specification of the time-dependent effect (Wald test of model: $p = 0.373$) does not fit as well as the linear one ($p = 0.011$). Although the LOWESS curve in Figure 1C would suggest that the linear time-dependent effect is not ideal, it still yields a better model fit than the log-linear time-dependent effect, which assumes a more rapid change shortly after start of follow-up. Time-dependent effects agreeing more closely with the LOWESS curve could be obtained by modeling more complex functions of time $\gamma(t)$ derived in a data-dependent way. For example, restricted cubic splines (Heinzel and Kaider 1997) could be applied using `coxph` with the `rcs` argument. However, in the gastric cancer study with its relatively small sample size the danger of overfitting the data outweighs the possible merits of a more complex modeling approach.

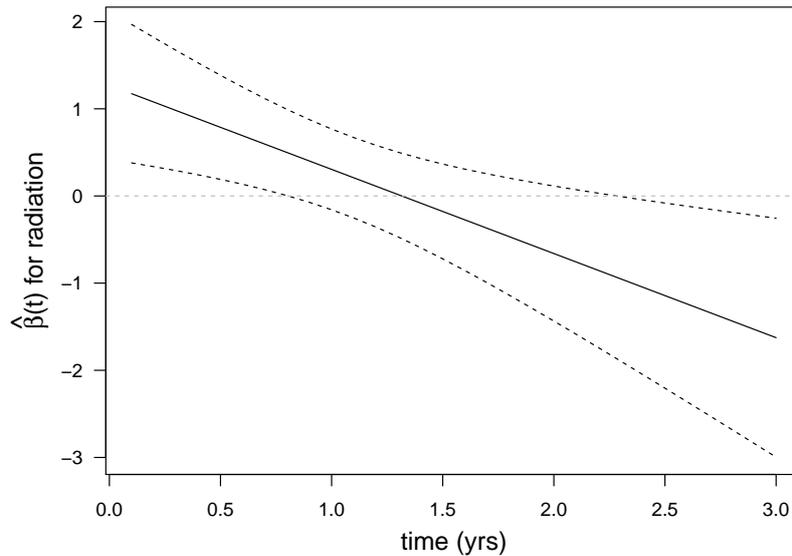


Figure 4: Gastric cancer study: Plot of the linear time-dependent effect of radiation with 95% confidence limits obtained by applying the plot method to an object of class ‘predict.coxphw’.

Weighted Cox regression

Weighted estimation of Cox regression gives a properly population-averaged estimate of the effect of radiation on survival also in case of non-proportional hazards. Using `coxphw` such a population-averaged HR can be estimated by the following statement:

```
R> fit2 <- coxphw(Surv(yrs, status) ~ radiation, data = gastric,
+   template = "AHR")
R> summary(fit2)
```

```
coxphw(formula = Surv(yrs, status) ~ radiation, data = gastric,
        template = "AHR")
```

Model fitted by weighted estimation (AHR template)

	coef	se(coef)	exp(coef)	lower 0.95	upper 0.95	z	p
radiation	0.463	0.239	1.588	0.995	2.536	1.937	0.053

Wald Chi-square = 3.753 on 1 df, p = 0.053

Generalized concordance probability:

	concordance prob.	lower 0.95	upper 0.95
radiation	0.6136	0.4986	0.7172

While standard Cox regression estimated the HR of radiation as 1.15 (95% CI: 0.74–1.81; $p = 0.537$), the average HR estimate by weighted estimation is 1.59 (95% CI: 1.00–2.54; $p = 0.053$). As explained above, from the average HR estimate the generalized concordance

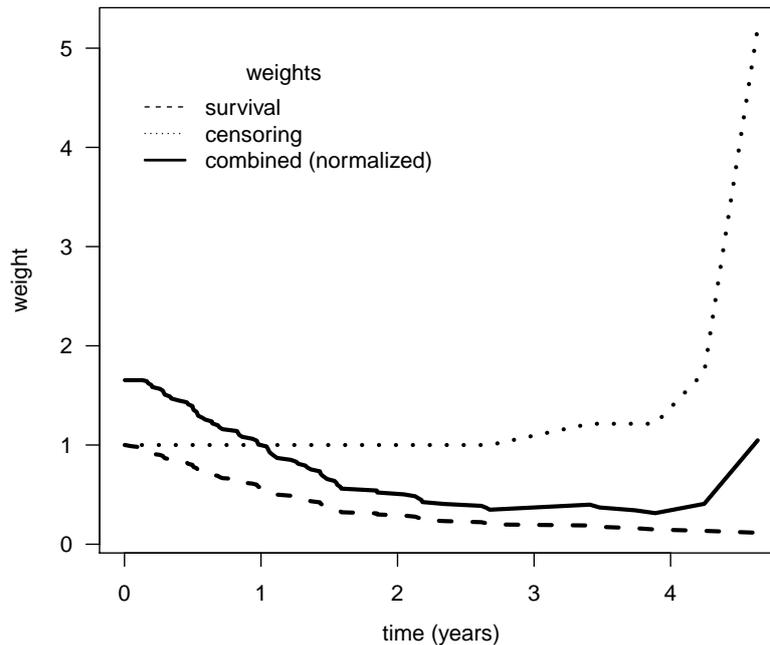


Figure 5: Gastric cancer study: Weights used by weighted Cox regression are plotted against time. Survival weights, $\hat{S}(t_j)$, are the left-continuous survivor function estimates at time t_j ; censoring weights, $\hat{G}(t_j)^{-1}$, are estimates of the follow-up distribution obtained by Kaplan-Meier estimation with reversed meaning of the status indicator at time t_j ; and combined (normalized) weights are the product of the survival and the censoring weights, rescaled to a mean of 1.

probability \hat{c}' can be derived, which in our example is computed as 0.6136. This means, that with a probability of 61.36% (95% CI: 49.86–71.72%), a patient treated with radiation is expected (at baseline) to die earlier than a patient treated without radiation.

If a few “long survivors” are present in a data set, weights used in weighted Cox regression may exhibit a very skew distribution. Such extremely unequal weighting will not bias the estimates, but may result in low precision of the estimates, i.e., wide confidence intervals. Truncation of weights at a suitable value, e.g., their 95th percentile, may greatly increase precision while introducing only a small amount of bias (Dunkler *et al.* 2010). The distribution of weights can be reviewed by using the `plot` method.

```
R> plot(fit2)
```

In this example (Figure 5), the censoring weights are very small during the first four years and increase sharply afterwards. However, the normalized total weights have a fairly narrow range from 0.31 to 1.74. Hence, truncation of weights using the `trunc.weights` argument `coxphw` may not seem necessary. If the weights were truncated at the, e.g., 95th percentile, the range of the normalized total weights would barely change to [0.31, 1.65], and this would yield very similar average HR estimates (HR 1.588; 95% CI: 0.995–2.533; $p = 0.053$) as without truncation of weights.

Table 2 summarizes the results of the discussed methods to analyze the time-dependent effect of radiation. A standard Cox regression would falsely give the impression that survival

	Estimate (95% CI)	<i>p</i> value
<i>Ignoring non-proportional hazards</i> *		
HR Cox regression	1.152 (0.735–1.805)	0.537
<i>Estimating piecewise constant HRs</i> *		
HR 1st year	2.405 (1.270–4.554)	0.007
HR >1st year	0.546 (0.277–1.078)	0.081
<i>Including a time-by-covariate interaction</i>		
HR at 0.5 years	2.197 (1.210–3.992)	0.010
HR at 1 year	1.356 (0.854–2.155)	0.197
HR at 2 years	0.517 (0.238–1.121)	0.095
<i>Weighted Cox regression</i>		
average HR	1.588 (0.995–2.533)	0.053
$\hat{c}'\%$	61.35 (49.87–71.70)	0.053

Table 2: Gastric cancer study: Four approaches to estimate the time-dependent effect of **radiation**. Estimates are either hazard ratios (HRs) or, if applicable, the generalized concordance probability in percent $\hat{c}'\%$ with their 95% confidence intervals (CI), and Wald test *p* values. In the model including a time-by-covariate interaction a linear time-dependent effect was assumed. * indicates possibly biased HR estimation based on models ignoring non-proportional hazards of **radiation**.

of a patient with and without radiation is similar. Modeling the time-dependent effect in more detail reveals clear changes over time. Average HR as an effect size measure and the generalized concordance probability \hat{c}' are suitable to support clear-cut treatment decisions. In our example, results point towards better prognosis under chemotherapy alone, but the difference in survival between the two groups fails to reach significance at the 5% level.

6.2. Biofeedback therapy study

In Section 2.2 we saw that biofeedback therapy had a time-dependent effect on full swallowing rehabilitation when adjusted for healing time. As the sign of the time-dependent log HR, $\hat{\beta}(t)$, changes over time (Figure 1B), the time-dependency of the effect of **bfb** should not be ignored by using a standard Cox regression. Applying it nonetheless, the HR of biofeedback therapy compared to standard therapy adjusted for **log2heal** is estimated as 1.31 (95% CI: 0.67–2.58; *p* = 0.434).

Since the sample is small and both variables **bfb** and **log2heal** exhibit non-proportional hazards neither stratification nor explicit modeling of the time-dependent effects are appropriate. [Denk and Kaider \(1997\)](#) restricted the follow-up time to 40 days and applied a Cox regression for **bfb** adjusted for **log2heal**. Now that weighted Cox regression is available, the data could be re-analyzed in order to derive the average HR of **bfb** adjusted for **log2heal** and a corresponding adjusted estimate of the generalized concordance probability \hat{c}' . Unlike the original analysis of [Denk and Kaider \(1997\)](#), this would greatly support decisions on the best therapy for patients to achieve fast swallowing rehabilitation after the healing process has been completed.

In this example, it is not just that both covariates exhibit time-dependent effects, but `log2heal` is also a strong confounder of `bfb`. Misspecification of the effect of one variable may thus lead to bias in the estimate of the other variable. For example, if we aim at estimating an average HR of `bfb` we are implicitly assuming a constant effect of that variable, which is not correct and hence may induce bias in the estimation of the regression coefficients of `log2heal` and its interaction with time. As a remedy, we propose the following two-stage estimation procedure which ensures that the effect estimate of `log2heal` is approximately unbiased and still allows for an unbiased assessment of the average HR of `bfb`:

Stage 1. The time-dependent effect of `log2heal` is estimated while appropriately taking into account the non-proportional hazards of `bfb`. One way to achieve this is to estimate an extended Cox regression model which includes the interaction of `log2heal` with time, and is stratified by `bfb`.

Stage 2. Weighted Cox regression is employed to estimate the average HR of `bfb` while holding the regression coefficients of `log2heal` and the interaction of `log2heal` with time fixed at their respective values estimated at stage 1.

The model of stage 1 can be estimated using `coxph`. Specifically, we estimate a log-linear time-dependent effect of `log2heal`, stratified for `bfb`. (A linear interaction with time was also evaluated, but did not improve the model fit.)

```
R> stage1 <- coxph(Surv(thdur, success) ~ strata(bfb) + log2heal +
+   tt(log2heal) + cluster(id), data = biofeedback,
+   tt = function(x, t, ...) x * log(t), method = "breslow")
```

```
coxph(formula = Surv(thdur, success) ~ strata(bfb) + log2heal +
+   tt(log2heal) + cluster(id), data = biofeedback,
+   tt = function(x, t, ...) x * log(t), method = "breslow")
```

```
n= 33, number of events= 23
```

	coef	exp(coef)	se(coef)	robust se	z	p
log2heal	0.737	2.089	0.901	0.380	1.940	0.052
tt(log2heal)	-0.415	0.660	0.326	0.149	-2.786	0.005

At stage 2 an average HR for `bfb` adjusted for the time-dependent effect of `log2heal` estimated at stage 1 is estimated using `coxphw`. Regression coefficients for the effects of `log2heal` and `tt(log2heal) = log(thdur):log2heal` can be held constant by making use of the `betafix` argument of `coxphw`.

```
R> summary(coxphw(Surv(thdur, success) ~ bfb + log2heal +
+   log(thdur):log2heal, data = biofeedback, template = "AHR",
+   betafix = c(NA, coef(stage1)))
```

This call to `coxphw`, setting the first element of `betafix` to `NA`, specifies that the first regression coefficient (`bfb`) should be estimated without restrictions. The values of the other two regression coefficients are held constant at the values extracted from the model of stage 1 using `coef(stage1)`. This call of `coxphw` will give the following output:

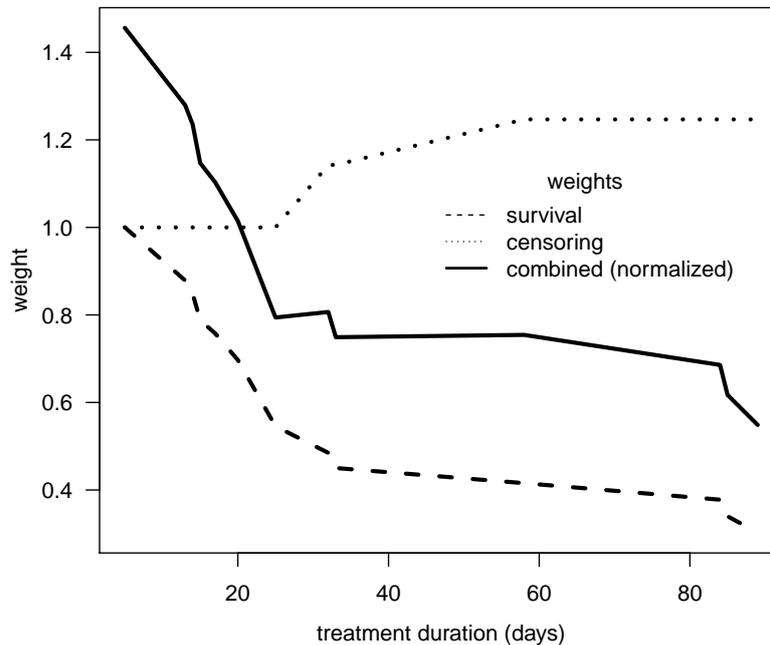


Figure 6: Biofeedback therapy study: Weights used by weighted Cox regression are plotted against time. For information on the different weights see Figure 5.

```
coxphw(formula = Surv(thdur, success) ~ bfb + log2heal + log(thdur):log2heal,
       data = biofeedback, template = "AHR", betafix = c(NA, coef(stage1)))
```

Model fitted by weighted estimation (AHR template)

	coef	se(coef)	exp(coef)	lower0.95	upper0.95	z	p
bfb	0.597	0.373	1.816	0.874	3.775	1.599	0.110
log2heal	0.737	NA	2.089	NA	NA	NA	NA
log(thdur):log2heal	-0.415	NA	0.660	NA	NA	NA	NA

Wald Chi-square = 2.556 on 1 df, $p = 0.110$ (based on: bfb)

Generalized concordance probability:

	concordance prob.	lower 0.95	upper 0.95
bfb	0.6449	0.4663	0.7906
log2heal	0.6763	NA	NA
log(thdur):log2heal	0.3977	NA	NA

Note that for `log2heal` and the interaction of `log2heal` with time, `log(thdur):log2heal`, no standard error estimates are given, as their regression coefficients were not re-estimated in the model. The global Wald test only relates to those variables for which regression coefficients were estimated, and thus, in our example, refers to `bfb` only.

The average HR of `bfb` when adjusted for `log2heal` is estimated as 1.82 (95% CI: 0.87–3.78; $p = 0.110$). From the generalized concordance probability \hat{c}' we can conclude that

among all (hypothetical) pairs of patients with equal healing time but with different therapies, therapy with biofeedback will earlier lead to full swallowing rehabilitation than therapy without biofeedback in 64.49% of the cases (95% CI: 46.63–79.06). Of course, while being an interesting finding, the result is failing to reach conventional levels of significance.

Weights applied in weighted Cox estimation can be checked using the `plot` function (Figure 6). We see that the variability of the normalized total weights is sufficiently small in this example.

7. Concluding remarks and summary

The choice of an adequate method to analyze survival data in the presence of non-proportional hazards will always depend on the type of research question. If researchers are interested in prediction of survival time, then they should explicitly model a time-dependent effect, probably in a data-dependent fashion, but observing general rules of statistical modeling (Harrell, Lee, and Mark 1996). However, for models to be used in etiologic research, i.e., to estimate causal effects of treatments or risk factors, a single number such as the average HR or the generalized concordance probability c' will often be more suitable and better interpretable. These measures allow non-parametric comparisons of life-expectancy between groups of subjects, which are also easier to communicate to non-statisticians.

The average HR (and the related generalized concordance probability c') can be estimated by weighted Cox regression using the weighting function $w(t) = S(t_j)G(t_j)^{-1}$ (Schemper *et al.* 2009). Other, earlier proposed weighting schemes do not necessarily lead to average HRs that can be transformed to concordance probabilities (Schemper 1992; Sasieni 1993; Xu and O'Quigley 2000). Unweighted Cox regression also provides a summary effect, but in case of non-proportional hazards its estimate is dependent on the censoring pattern. Since censoring is a property of a study and usually not of a population, estimates from standard Cox regression are not generalizable if censoring and time-dependence of effects are substantial. In case of proportional hazards, (unweighted) Cox regression will provide the most efficient estimates, but weighted Cox regression can still be safely applied as it will not introduce bias.

In Tables 3 and 4 we summarize methods of analysis in the presence of non-proportional hazards. In this paper, we introduced the R package `coxphw` implementing weighted Cox regression. This package significantly extends the R toolbox for survival analysis. Beyond the estimation of static weighted Cox regression models it enables the inclusion of interactions of covariates with functions of time for unweighted and weighted Cox regression models by using an intuitive syntax in the `formula` argument. Regression coefficients of covariates can be fixed at specified values, even if these covariates are assumed to have time-dependent effects. As exemplified in the biofeedback therapy study, this allows to implement a two-stage estimation process in which an average hazard ratio is estimated for one covariate while accounting for a time-dependent effect of another covariate.

The R package `coxphw` is available from the Comprehensive R Archive Network (CRAN) at <https://CRAN.R-project.org/package=coxphw> or at <http://cemsii.meduniwien.ac.at/en/kb/science-research/software/>. The latter link also points to a SAS (SAS Institute Inc. 2013) macro `WCM` enabling weighted estimation of Cox regression.

Method	Interpretation of $\hat{\beta}$	Additional DFs	Functions in R
<i>Ignoring non-PH: Cox regression</i>	log HR	0	<code>coxph</code> <code>coxphw</code>
<i>Weighted Cox regression with $\hat{w}(t_j) = \hat{G}(t_j)^{-1}$ (Xu and O'Quigley 2000)</i>	average regression effect	0	<code>coxphw</code>
<i>Weighted Cox regression with $\hat{w}(t_j) = \hat{S}(t_j)\hat{G}(t_j)^{-1}$ (Schemper <i>et al.</i> 2009)</i>	log average HR	0	<code>coxphw</code>
Concordance regression (Dunkler <i>et al.</i> 2010)	log odds of concordance	0	<code>concreg</code>
<i>Piecewise constant HRs</i>	piecewise log HR	1 for each additional period	<code>coxph</code> <code>coxphw</code>
<i>Stratification</i>	no explicit estimation, only adjustment	0	<code>coxph</code> with <code>strata</code>
<i>Including time-by-covariate interactions: Explicit specification (e.g., linear, log-linear)</i>	time-dependent log HR	dependent on function (linear & log-linear: 1)	<code>coxph</code> , <code>coxphw</code>
<i>Including time-by-covariate interactions: Data-driven specification (e.g., RCS)</i>	time-dependent log HR	RCS: number of knots-1	RCS: <code>coxph</code>

Table 3: Various methods of analysis of survival data with non-proportional hazards in R. An appropriate analysis may depend on prior knowledge of possible time-dependent effects, the aims of the analysis, the sample size, and number of events. `coxph` is implemented in the package `survival` (Therneau 2017); `coxphw` in `coxphw`; and `concreg` in `concreg` (Heinze *et al.* 2016). Package `rms` (Harrell 2017) has to be loaded before applying restricted cubic splines using `coxph`. DF, degrees of freedom; $\hat{G}(t_j)$, estimate of the follow-up distribution obtained by Kaplan-Meier estimation with reversed meaning of the status indicator at time t_j ; HR, hazard ratio; PH, proportional hazards; RCS, restricted cubic splines; $\hat{S}(t_j)$, left-continuous Kaplan-Meier estimate of the survivor function at time t_j .

Method	Advantages & disadvantages
<i>Ignoring non-PH:</i> Cox regression	+ simple and efficient, if HR is always < 1 or > 1 – estimates depending on type of non-PH and on censoring pattern
<i>Weighted Cox regression</i> with $\hat{w}(t_j) = \hat{G}(t_j)^{-1}$ (Xu and O’Quigley 2000)	+ provides an average effect that is independent of the observed censoring pattern – “average regression effect” is not interpretable as log odds of concordance – “average regression effect” is not suitable for prediction
<i>Weighted Cox regression</i> with $\hat{w}(t_j) = \hat{S}(t_j)\hat{G}(t_j)^{-1}$ (Schemper <i>et al.</i> 2009)	+ independent from type of non-PH and censoring pattern + good approximation of c' for HRs up to a value of 4 – concept of average HRs is not suitable for prediction
Concordance regression (Dunkler <i>et al.</i> 2010)	+ independent from type of non-PH and censoring pattern + estimates c' – less efficient than weighted Cox regression with $\hat{w}(t_j) = \hat{S}(t_j)\hat{G}(t_j)^{-1}$
<i>Piecewise constant HRs</i>	+ simple interpretation – usually only rough approximation of the true time-dependent effect
<i>Stratification</i>	+ no need to specify the type of time-dependency – only suitable for a categorical covariate which is not of primary interest
<i>Including time-by-covariate interactions:</i> Explicit specification (e.g., linear, log-linear)	+ often suitable for monotone patterns of time-dependencies – not flexible
<i>Including time-by-covariate interactions:</i> Data-driven specification (e.g., RCS)	+ allows parsimonious, but flexible estimation of time-dependent effects – requires sufficient sample size and number of events – results may differ between alternative methods

Table 4: Advantages and disadvantages of various methods to analyze survival data with non-proportional hazards. c' , generalized concordance probability; $\hat{G}(t_j)$, estimate of the follow-up distribution obtained by Kaplan-Meier estimation with reversed meaning of the status indicator at time t_j ; HR, hazard ratio; PH, proportional hazards; RCS, restricted cubic splines; $\hat{S}(t_j)$, left-continuous Kaplan-Meier estimate of the survivor function at time t_j .

Acknowledgments

This work was partly funded by the European Union (Grant Agreement No. HEALTH F2-2009-241544). We are grateful to Maria Kohl, Medical University of Vienna, for technical help and Dr. D.-M. Denk, Medical University of Vienna, for providing the data of the biofeedback therapy study.

References

- Bender R, Augustin T, Blettner M (2005). “Generating Survival Times to Simulate Cox Proportional Hazards Models.” *Statistics in Medicine*, **24**(11), 1713–1723. doi:10.1002/sim.2059.
- Breslow N (1970). “A Generalized Kruskal-Wallis Test for Comparing K Samples Subject to Unequal Patterns of Censorship.” *Biometrika*, **57**(3), 579–594. doi:10.2307/2334776.
- Breslow N, Crowley J (1974). “A Large Sample Study of the Life Table and Product Limit Estimates under Random Censorship.” *The Annals of Statistics*, **2**(3), 437–453. doi:10.1214/aos/1176342705.
- Cox DR (1972). “Regression Models and Life-Tables.” *Journal of the Royal Statistical Society B*, **34**(2), 187–220.
- Denk DM, Kaider A (1997). “Videoendoscopic Biofeedback: A Simple Method to Improve the Efficacy of Swallowing Rehabilitation of Patients After Head and Neck Surgery.” *Journal for Oto-Rhino-Laryngology, Head and Neck Surgery*, **59**(2), 100–105. doi:10.1159/000276918.
- Dunkler D, Schemper M, Heinze G (2010). “Gene Selection in Microarray Survival Studies under Possibly Non-Proportional Hazards.” *Bioinformatics*, **26**(6), 784–790. doi:10.1093/bioinformatics/btq035.
- Gastrointestinal Tumor Study Group (1982). “A Comparison of Combination Chemotherapy and Combined Modality Therapy for Locally Advanced Gastric Carcinoma.” *Cancer*, **49**(9), 1771–1777. doi:10.1002/1097-0142(19820501)49:9<1771::AID-CNCR2820490907>3.0.CO;2-M.
- Grambsch PM, Therneau TM (1994). “Proportional Hazards Tests and Diagnostics Based on Weighted Residuals.” *Biometrika*, **81**(3), 515–526. doi:10.1093/biomet/81.3.515.
- Hanley JA, McNeil BJ (1982). “The Meaning and Use of the Area under a Receiver Operating Characteristic (ROC) Curve.” *Radiology*, **143**(1), 29–36. doi:10.1148/radiology.143.1.7063747.
- Harrell FE (2001). *Regression Modeling Strategies. With Applications to Linear Models, Logistic Regression, and Survival Analysis*. Springer-Verlag, New York.
- Harrell FE (2017). *rms: Regression Modeling Strategies*. R package version 5.1-1, URL <https://CRAN.R-project.org/package=rms>.

- Harrell FE, Lee KL, Mark DB (1996). “Multivariable Prognostic Models: Issues in Developing Models, Evaluating Assumptions and Adequacy, and Measuring and Reducing Errors.” *Statistics in Medicine*, **15**(4), 361–387. doi:10.1002/(sici)1097-0258(19960229)15:4<361::aid-sim168>3.0.co;2-4.
- Heinze G, Ploner M, Dunkler D (2016). **conreg**: *Concordance Regression*. R package version 0.6, URL <https://CRAN.R-project.org/package=conreg>.
- Heinze G, Ploner M, Dunkler D (2018). **coxphw**: *Weighted Estimation in Cox Regression*. R package version 4.0.1, URL <https://CRAN.R-project.org/package=coxphw>.
- Heinzel H, Kaider A (1997). “Gaining More Flexibility in Cox Proportional Hazards Regression Models with Cubic Spline Functions.” *Computer Methods and Programs in Biomedicine*, **54**(3), 201–208. doi:10.1016/s0169-2607(97)00043-6.
- Lehr S, Schemper M (2007). “Parsimonious Analysis of Time-Dependent Effects in the Cox Model.” *Statistics in Medicine*, **26**(13), 2686–2698. doi:10.1002/sim.2742.
- Lin DY (1991). “Goodness-of-Fit Analysis for the Cox Regression Model Based on a Class of Parameter Estimators.” *Journal of the American Statistical Association*, **86**(415), 1074–1078. doi:10.2307/2290404.
- Lin DY, Wei LJ (1989). “The Robust Inference for the Cox Proportional Hazards Model.” *Journal of the American Statistical Association*, **84**(408), 1074–1078. doi:10.1080/01621459.1989.10478874.
- Mann HB, Whitney DR (1947). “On a Test of Whether One of 2 Random Variables Is Stochastically Larger Than the Other.” *The Annals of Mathematical Statistics*, **18**(1), 50–60. doi:10.1214/aoms/1177730491.
- Mantel N (1966). “Evaluation of Survival Data and Two New Rank Order Statistics Arising in Its Consideration.” *Cancer Chemother Reports*, **50**(3), 163–170.
- Prentice RL (1978). “Linear Rank Tests with Right Censored Data.” *Biometrika*, **65**(1), 167–179. doi:10.2307/2335292.
- R Core Team (2017). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- Royston P, Altman DG (1994). “Regression Using Fractional Polynomials of Continuous Covariates: Parsimonious Parametric Modelling.” *Journal of the Royal Statistical Society C*, **43**(3), 429–467. doi:10.2307/2986270.
- Sasieni P (1993). “Maximum Weighted Partial Likelihood Estimators for the Cox Model.” *Journal of the American Statistical Association*, **88**(421), 144–152. doi:10.1080/01621459.1993.10594304.
- SAS Institute Inc (2013). *The SAS System, Version 9.4*. SAS Institute Inc., Cary. URL <http://www.sas.com/>.
- Schemper M (1992). “Cox Analysis of Survival Data with Non-Proportional Hazard Functions.” *The Statistician*, **41**, 455–465. doi:10.2307/2349009.

- Schemper M, Smith TL (1996). “A Note on Quantifying Follow-Up in Studies of Failure Time.” *Controlled Clinical Trials*, **17**(4), 343–346. doi:10.1016/0197-2456(96)00075-x.
- Schemper M, Wakounig S, Heinze G (2009). “The Estimation of Average Hazard Ratios by Weighted Cox Regression.” *Statistics in Medicine*, **28**(19), 2473–2489. doi:10.1002/sim.3623.
- Steyerberg EW, Eijkemans MJC, Harrell FE, Habbema JDF (2000). “Prognostic Modelling with Logistic Regression Analysis: A Comparison of Selection and Estimation Methods in Small Data Sets.” *Statistics in Medicine*, **19**(8), 1059–1079. doi:10.1002/(SICI)1097-0258(20000430)19:8<1059::AID-SIM412>3.0.CO;2-0.
- Therneau TM (2017). *survival: A Package for Survival Analysis in S*. R package version 2.41-3, URL <https://CRAN.R-project.org/package=survival>.
- Therneau TM, Grambsch PM (2000). *Modeling Survival Data: Extending the Cox Model*. Springer-Verlag, New York.
- Thisted RA (1988). *Elements of Statistical Computing: Numerical Computation*. CRC Press, New York.
- Wakounig S, Heinze G, Schemper M (2015). “Non-Parametric Estimation of Relative Risk in Survival and Associated Tests.” *Statistical Methods in Medical Research*, **24**(6), 856–870. doi:10.1177/0962280211431022.
- Xu R, O’Quigley J (2000). “Estimating Average Regression Effect under Non-Proportional Hazards.” *Biostatistics*, **1**(4), 423–439. doi:10.1093/biostatistics/1.4.423.

Affiliation:

Georg Heinze
Section for Clinical Biometrics
Center for Medical Statistics, Informatics and Intelligent Systems
Medical University of Vienna
1090 Vienna, Austria
E-mail: georg.heinze@meduniwien.ac.at
URL: <http://cemsiiis.meduniwien.ac.at/en/kb>