Macro Risk_Est_PH_reg_time_dep User Guide

Michael Crager
9 May 2019

This macro calculates point and interval estimates of the risk of an event using Cox proportional hazards regression. “External” time-dependent covariates, stratified cohort sampling study designs, and left truncation are accommodated.

Statistical Methods

Suppose we have a multivariate Cox (1972) proportional hazards model with estimated parameter vector $\hat{\boldsymbol{\beta}} = \left(\hat{\beta}_1, \hat{\beta}_2, \ldots, \hat{\beta}_p\right)^{\mathrm{T}}$ and observed covariate vectors $\mathbf{z}_i = \left(z_{i1}, z_{i2}, \ldots, z_{ip}\right)^{\mathrm{T}}$, $i = 1, 2, \ldots, n$, where $Y_i(t)$ is the indicator of whether patient i is still in the risk set at time *t*. Let $\bar{N}(t) = \sum_{i=1}^{n} N_i(t)$ be the counting process giving the total number of patients who have reached an endpoint by time *t*, where $N_i(t)$ is the counting process for patient i .

The Breslow (1972) estimate of the baseline cumulative hazard function $\Lambda_0(T)$ is

$$\hat{\Lambda}_0(T) = \int_0^T \frac{\mathrm{d}\,\bar{N}(t)}{\sum_{i=1}^{n} Y_i(t) \exp\left(\hat{\boldsymbol{\beta}}^{\mathrm{T}} \mathbf{z}_i\right)}$$

For a patient with covariate vector $\mathbf{z}$, the estimated cumulative hazard function is

$$\hat{\Lambda}(T; \mathbf{z}) = \hat{\Lambda}_0(T) \exp\left(\hat{\boldsymbol{\beta}}^{\mathrm{T}} \mathbf{z}\right)$$

A consistent estimate of the variance of the baseline cumulative hazard (see Therneau and Grambsch, 2000) is

$$\widehat{\mathrm{Var}}\left\{\hat{\Lambda}_0(T)\right\} = \int_0^T \frac{\mathrm{d}\,\bar{N}(t)}{\left\{\sum_{i=1}^{n_k} Y_i(s) \exp\left(\boldsymbol{\beta}^{\mathrm{T}} \mathbf{z}_i\right)\right\}^2} \tag{1}$$

Tsiatis’s (1981) estimate of the variance of the estimated cumulative hazard $\hat{\Lambda}(T; \mathbf{z})$ for an individual patient with covariate vector $\mathbf{z}$, is

$$\widehat{\mathrm{Var}}\left\{\hat{\Lambda}(T; \mathbf{z})\right\} = \exp\left(2\hat{\boldsymbol{\beta}}^{\mathrm{T}} \mathbf{z}\right) \widehat{\mathrm{Var}}\left\{\hat{\Lambda}_0(T)\right\} + \mathbf{q}^{\mathrm{T}} \hat{\mathbf{V}} \mathbf{q} \tag{2}$$

where

$$\mathbf{q} = \exp\left(\hat{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{z}\right)\int_0^T \left\{\mathbf{z} - \overline{\mathbf{z}}(t)\right\} \frac{\mathrm{d}\,\overline{N}(t)}{\sum_{i=1}^n Y_i(s)\exp\left(\hat{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{z}_i\right)}$$

and

$$\overline{\mathbf{z}}(t) = \frac{\sum_{i=1}^n \mathbf{z}_i Y_i(t)\exp\left(\hat{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{z}_i\right)}{\sum_{i=1}^n Y_i(t)\exp\left(\hat{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{z}_i\right)}$$

For a cohort sampling study where patient i has sampling weight w_i, $i = 1, 2, \ldots, n$, we fit the regression parameters $\boldsymbol{\beta}$ by maximizing the pseudolikelihood and estimate the baseline cumulative hazard function as

$$\hat{\Lambda}_0^{(w)}(T) = \int_0^T \frac{\sum_{i=1}^n w_i \,\mathrm{d}\,N_i(t)}{\sum_{i=1}^n w_i Y_i(t)\exp\left(\hat{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{z}_i\right)} \tag{3}$$

Its variance is consistently estimated by

$$\widehat{\mathrm{Var}}\left(\hat{\Lambda}_0^{(w)}(T)\right) = \int_0^T \frac{\sum_{i=1}^n w_i^2 \,\mathrm{d}\,N_i(t)}{\left\{\sum_{i=1}^n w_i Y_i(t)\exp\left(\hat{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{z}_i\right)\right\}^2}$$

(see Appendix 1). Tsiatis's (1981) estimate of the variance of the estimated cumulative hazard $\hat{\Lambda}(T;\mathbf{z})$ for an individual patient with covariate vector $\mathbf{z}$ becomes

$$\widehat{\mathrm{Var}}\left\{\hat{\Lambda}^{(w)}(T;\mathbf{z})\right\} = \exp\left(2\hat{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{z}\right)\widehat{\mathrm{Var}}\left\{\hat{\Lambda}_0^{(w)}(T)\right\} + \mathbf{q}^{\mathrm{T}}\hat{\mathbf{V}}\mathbf{q} \tag{4}$$

where the covariance matrix $\hat{\mathbf{V}}$ can be estimated using the method of Lin and Wei (1989), and

$$\mathbf{q} = \exp\left(\hat{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{z}\right)\int_0^T \left\{\mathbf{z} - \overline{\mathbf{z}}(t)\right\} \frac{\sum_{i=1}^n w_i \,\mathrm{d}\,N_i(t)}{\sum_{i=1}^n w_i Y_i(s)\exp\left(\hat{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{z}_i\right)}$$

$$\overline{\mathbf{z}}(t) = \frac{\sum_{i=1}^n w_i \mathbf{z}_i Y_i(t)\exp\left(\hat{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{z}_i\right)}{\sum_{i=1}^n w_i Y_i(t)\exp\left(\hat{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{z}_i\right)}$$

Finally, if the covariates are "externally" time-dependent, that is, $\mathbf{z}$ is replaced by the function $\mathbf{z}(t)$, where the relationship of the covariate vector with time is fixed given the regression parameters, the estimated cumulative hazard at time T for a patient with covariate vector path $\mathbf{z}(\bullet)$ is

$$\hat{\Lambda}^{(w)}(T;\mathbf{z}(\bullet)) = \int_0^T \frac{\exp\left(\hat{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{z}(t)\right)\sum_{i=1}^n w_i dN_i(t)}{\sum_{i=1}^n w_i Y_i(t)\exp\left(\hat{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{z}_i(t)\right)} \tag{5}$$

and the corresponding consistent estimate of variance is

$$\widehat{\mathrm{Var}}\left(\hat{\Lambda}^{(w)}(T;\mathbf{z}(\bullet))\right) = \int_0^T \frac{\exp\left(2\hat{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{z}(t)\right)\sum_{i=1}^n w_i^2 \,\mathrm{d}\, N_i(t)}{\left\{\sum_{i=1}^n w_i Y_i(t)\exp\left(\hat{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{z}_i(t)\right)\right\}^2} + \mathbf{q}^{\mathrm{T}}\hat{\mathbf{V}}\mathbf{q} \tag{6}$$

with

$$\mathbf{q} = \int_0^T \left\{\mathbf{z}(t) - \overline{\mathbf{z}}(t)\right\} \frac{\exp\left(\hat{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{z}(t)\right)\sum_{i=1}^n w_i \,\mathrm{d}\, N_i(t)}{\sum_{i=1}^n w_i Y_i(s)\exp\left(\hat{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{z}_i(t)\right)}$$

$$\overline{\mathbf{z}}(t) = \frac{\sum_{i=1}^n w_i \mathbf{z}_i(t) Y_i(t)\exp\left(\hat{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{z}_i(t)\right)}{\sum_{i=1}^n w_i Y_i(t)\exp\left(\hat{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{z}_i(t)\right)}$$

If we set $\delta_i = 1$ if patient i had an event and 0 if not, let t_i denote the event time for patient i , and let $\mathbb{R}(t) = \left\{ j : t_j \geq t \right\}$ denote the set of patients still at risk at time t, we can rewrite (5) as

$$\hat{\Lambda}^{(w)}\left(T;\mathbf{z}(\bullet)\right) = \sum_{i:\delta_i = 1, t_i \leq T} \left\{ \frac{w_i \exp\left(\hat{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{z}(t_i)\right)}{\sum_{j \in \mathbb{R}(t_i)} w_j \exp\left(\hat{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{z}_j(t_i)\right)} \right\} \tag{7}$$

and expression (6) as

$$\widehat{\mathrm{Var}}\left(\hat{\Lambda}^{(w)}(T;\mathbf{z}(\bullet))\right) = \sum_{i:\delta_i = 1, t_i \leq T} \left[\frac{w_i^2 \exp\left(2\hat{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{z}(t_i)\right)}{\left\{\sum_{j \in \mathbb{R}(t_i)} w_j \exp\left(\hat{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{z}_j(t_i)\right)\right\}^2} \right] + \mathbf{q}^{\mathrm{T}}\hat{\mathbf{V}}\mathbf{q} \tag{8}$$

where

$$\mathbf{q} = \sum_{i:\delta_i=1, t_i \le T} \left(\mathbf{z}(t_i) - \overline{\mathbf{z}}(t_i)\right) \frac{w_i \exp\left(\hat{\boldsymbol{\beta}}^{\mathrm{T}} \mathbf{z}(t_i)\right)}{\sum_{j \in \mathbb{R}(t_i)} w_j \exp\left(\hat{\boldsymbol{\beta}}^{\mathrm{T}} \mathbf{z}_j(t_i)\right)}$$

$$\overline{\mathbf{z}}(t_i) = \frac{\sum_{j \in \mathbb{R}(t_i)} w_j \mathbf{z}_j(t_i) \exp\left(\hat{\boldsymbol{\beta}}^{\mathrm{T}} \mathbf{z}_j(t_i)\right)}{\sum_{j \in \mathbb{R}(t_i)} w_j \exp\left(\hat{\boldsymbol{\beta}}^{\mathrm{T}} \mathbf{z}_j(t_i)\right)}$$

Note that the estimated cumulative hazard and its variance estimate can be computed from expression (5) and (6) [or equivalently, (7) and (8)] for studies that do not use cohort sampling by $w_i \equiv 1/n$ for $i = 1, 2, \ldots, n$ and for analyses that do not use time-dependent covariates by setting the covariate functions to be constants, that is, $\mathbf{z}(t) \equiv \mathbf{z}$ for all *t*.

Let $\rho(T)$ be the natural logarithm of the true cumulative hazard at time *T*, estimated as

$$\hat{\rho}(T) = \ln\left(\hat{\Lambda}^{(w)}(T)\right)$$

Using the delta method, the standard error of $\hat{\rho}(T)$ is consistently estimated by

$$\widehat{\mathrm{SD}}\left\{\hat{\rho}(T)\right\} = \sqrt{\widehat{\mathrm{Var}}\left(\hat{\Lambda}^{(w)}(T)\right)} \Big/ \hat{\Lambda}^{(w)}(T)$$

Thus we can compute the estimated risk of an event occurring by time *T* as

$$\hat{r}(T) = 1 - \exp\left\{-\exp\left(\hat{\rho}(T)\right)\right\} \tag{9}$$

with level α confidence interval

$$\begin{aligned}\Bigg(1 - \exp\left[-\exp\left\{\hat{\rho}(T) - \Phi^{-1}\left(1 - \frac{\alpha}{2}\right)\widehat{\mathrm{SD}}\left(\hat{\rho}(T)\right)\right\}\right], \\ 1 - \exp\left[-\exp\left\{\hat{\rho}(T) + \Phi^{-1}\left(1 - \frac{\alpha}{2}\right)\widehat{\mathrm{SD}}\left(\hat{\rho}(T)\right)\right\}\right]\Bigg)\end{aligned} \tag{10}$$

where Φ^{-1} is the inverse cumulative distribution function of the standard normal distribution. This is the “loglog” method calculation. Alternatively, the “log” method confidence interval

$$\begin{aligned}\Bigg(1 - \exp\left[-\hat{\Lambda}(T) + \Phi^{-1}\left(1 - \frac{\alpha}{2}\right)\widehat{\mathrm{SD}}\left(\hat{\Lambda}(T)\right)\right], \\ 1 - \exp\left[-\hat{\Lambda}(T) - \Phi^{-1}\left(1 - \frac{\alpha}{2}\right)\widehat{\mathrm{SD}}\left(\hat{\Lambda}(T)\right)\right]\Bigg)\end{aligned}$$

or the “linear” method confidence interval

$$\left(\hat{r}(T)-\exp\left[-\hat{\Lambda}(T)\right]\Phi^{-1}\left(1-\frac{\alpha}{2}\right)\widehat{\mathrm{SD}}\left(\hat{\Lambda}(T)\right),\right.$$
$$\left.\hat{r}(T)+\exp\left[-\hat{\Lambda}(T)\right]\Phi^{-1}\left(1-\frac{\alpha}{2}\right)\widehat{\mathrm{SD}}\left(\hat{\Lambda}(T)\right)\right)$$

may be used. The linear method may give confidence interval limits that lies outside the unit interval.

Ties in the event times can be handled by the Efron approach. In equation (7), if *k* patients (for convenience, denote them as patients $i=1,2,\ldots,k$) have events at exactly the same time *t*, then we replace each of their terms $w_i \exp\left(\hat{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{z}(t)\right)$ in the numerator of the summand in (7) by $\overline{w}\exp\left(\hat{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{z}(t)\right)$ where $\overline{w}=\sum_{i=1}^{k} w_i \big/ k$. In the denominator we replace each of the $w_j \exp\left(\hat{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{z}_j(t)\right),\ j=1,2,\ldots,k$ by their average $\sum_{j=1}^{k} w_j \exp\left(\hat{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{z}_j(t)\right)\Big/ k$.

The risk calculation is the same in case the observations are left-truncated, that is, study subjects do not enter the risk set until a subject-specific time. In this case, the indicator $Y_i(t)$ for membership in the risk set is 0 until the patient enters the study and then becomes 1 until the patient is no longer being followed. Analyses with left-truncation are useful for estimating the risk of one event conditional on occurrence of another (in which case the subset of patients experiencing the first event is analyzed with left truncation at the time the first event occurs). Left truncation can also be used to change the time scale of the analysis. For example, as an alternative to analyzing time on study, patient age could be used as the time scale, with left truncation at the age at which each patient entered the study.

Macro Risk_Est_PH_reg_time_dep

Macro Risk_Est_PH_reg_time_dep computes risk estimates from a proportional hazards regression using the methods described above. The macro is called as follows:

```
%Risk_Est_PH_reg_time_dep(
    /* Input Specification */  indsn=, byvar= ,vars=,
                              time=, censor=, censorlist=,
                              entrytime=,weight=,
                              programming_statements=%str(), calc_vars=, covariate_dsn=,
    /* Analysis Parameters */  risk_time=, robust=, print=, alpha=, strata=, CI_method=
    /* Output Specification */ outdsn=, Risk=, Risk_LCL=, Risk_UCL=,
                               CumHaz, CumHaz_LCL=,CumHaz_UCL=,
                               LogCumHaz=, SE_LogCumHaz=
             );
```

The macro parameters are defined in Table 1. The time dependence of the covariates, if any, is defined using programming statements (as are used for PROC PHREG). It is assumed that these programming statements, when applied to the input data set (and the covariate data set, if specified), will uniquely determine the covariate values at the time specified by the input data set variable given by the macro parameter time. Note the programming statements will need to be enclosed in %str() so that the semicolon(s) will not cause a syntax error.

Table 1. Macro Risk_Est_PH_reg_time_dep Parameters

Parameter	Type	Required?	Default Value	Description
indsn	$	Yes	(at temporary library)	(Libname reference and) file name containing input data set.
byvar	$	No	—	Optional list of variables to do the analysis by.
vars	$	Yes	—	List of input data set variables to be used as the covariate in the Cox model used to estimate the risk. If the programming statements create the variables that are to be included in the model, list the variables thus created.
time	#	Yes	—	Input data set variable containing the time to event (or censoring).
censor	#	Yes	—	Input data set variable indicating whether the observed time to event was censored.
censorlist	#	No	0	List of values of variable censor that indicate a censored observation. Default is the single value 0.
entrytime	#	No	—	Optional input data set variable containing a left truncation time for each observation.
weight	#	No	—	Input data set variable giving the observation's weight in the analysis. If this parameter is set, it is assumed that cohort sampling was used and resulted in the specified weights.

Table 1. Macro Risk_Est_PH_reg_time_dep Parameters

Parameter	Type	Required?	Default Value	Description
programming_ statements	$	No	—	%str()-enclosed text string including programming statements that will be inserted into proc PHREG and various data steps to compute the time-dependent covariate values. For example: programming_statements = %str(if time <= 3 then x_3 = 0; else x_3 = x;) If no programming statements are entered, the risk calculations will be made for covariates that are constant over time.
calc_vars	$	Yes, if time-dependent covariates are used	—	List of variables that are used in the calculation of the time-dependent covariate values. Include all non-time-dependent covariate values in this list, too. Leave this parameter blank if time dependent variables are not used.
covariate_dsn	$	No	(input data set)	(Libname reference and) the name of a data set that contains the covariate values for which the risk is to be estimated. The data set must have all the variables included in the model, or that are required to derive these variables if the model has time-dependent covariates derived using programming statements. The data set must also include the stratification variable if the model is stratified. If no covariate data set is specified, the risk will be estimated for every patient in the main input data set.
risk_time	#	Yes	—	This is the time at which the risk is assessed for each patient. That is, the risk is defined as the probability that the patient will have the event on or before risk_time.
print	$	No	yes	If this parameter is set to no, the PROC PHREG output will not be printed.
alpha	#	No	0.05	The macro will compute a 100(1-alpha)% confidence interval for the risk and cumulative hazard.
strata	$/#	No	—	Character string giving input data set variable by which the proportional hazards regression analysis will be stratified.
CI_method	$	No	loglog	Character string giving the method for computing the confidence intervals. If linear is specified, the confidence interval is computed on the risk scale. If log is specified, the confidence interval is computed on the cumulative hazard scale and transformed to the risk scale. If loglog is specified, the confidence interval is computed on the log cumulative hazard scale and transformed to the risk scale.

Table 1. Macro Risk_Est_PH_reg_time_dep Parameters

Parameter	Type	Required?	Default Value	Description
outdsn	$	Yes	(at temporary library)	(Libname reference and) the output data set name. This data set will contain all the records and variables of the covariate data set (or the input data set if no separate coavariate data set is specified) plus the variables named by the following eight macro parameters.
Risk	#	No	Risk	Name of output data set variable that will contain the risk estimate.
Risk_LCL	#	No	Risk_LCL	Name of output data set variable will contain the lower limit of a 1-alpha confidence interval for the risk.
Risk_ULCL	#	No	Risk_UCL	Name of output data set variable will contain the upper limit of a 1-alpha confidence interval for the risk.
CumHaz	#	No	CumHaz	Name of output data set variable that will contain the cumulative hazard estimate.
CumHaz_LCL	#	No	CumHaz_LCL	Name of output data set variable will contain the lower limit of a 1-alpha confidence interval for the cumulative hazard.
CumHaz_UCL	#	No	CumHaz_UCL	Name of output data set variable will contain the upper limit of a 1-alpha confidence interval for the cumulative hazard.
LogCumHaz	#	No	LogCumHaz	Name of output data set variable that will contain the log cumulative hazard estimate.
SE_LogCumHaz	#	No	SE_LogCumHaz	Name of output data set variable that will contain the estimated standard error of the log cumulative hazard estimate.

Appendix 1. Justification for the Estimate of the Variance of the Cumulative Hazard Estimator When Cohort Sampling Is Used

If we have follow-up times t_i for $i = 1, 2, \ldots, n$ patients and indicators δ_i of whether an event occurred, and denote the event count at time *t* by

$$\bar{N}(t) = \sum_{i=1}^{n} N_i(t) = \sum_{i:t_i \le t} \delta_i$$

and the number of patients in the risk set at time *t* by

$$\bar{Y}(t) = \sum_{i=1}^{n} I_{\{t_i \le t\}}$$

Without cohort sampling, the Breslow estimate of the cumulative hazard at time *T* is

$$\hat{\Lambda}(T) = \int_0^T \frac{d\bar{N}(t)}{\bar{Y}(t)}$$

and its estimated variance is

$$\widehat{\mathrm{Var}}\left(\hat{\Lambda}(T)\right) = \int_0^T \frac{d\bar{N}(t)}{\left\{\bar{Y}(t)\right\}^2}$$

Following Therneau and Grambsch (2000), $\Delta_h \bar{N}(t) = \bar{N}(t+h) - \bar{N}(t)$, that is, the number of events in the interval $(t, t+h)$, is approximately Poisson distributed for small *h*. Since there are $\bar{Y}(t)$ patients at risk at time *t*, $\Delta_h \bar{N}(t)$ is Poisson with mean

$$\int_t^{t+h} \bar{Y}(s)\lambda(s)ds \approx \bar{Y}(t)\lambda(t)h$$

where $\lambda(t)$ is the true hazard function. Now the variance of a Poisson variable equals its mean, so

$$\mathrm{Var}\left\{\Delta_h \bar{N}(t)\right\} = \mathrm{E}\left\{\Delta_h \bar{N}(t)\right\} \approx \bar{Y}(t)\lambda(t)h$$

and thus

$$\mathrm{Var}\left\{\Delta_h \bar{N}(t) / \bar{Y}(t)\right\} \approx \bar{Y}(t)\lambda(t)h \Big/ \left\{\bar{Y}(t)\right\}^2 = \lambda(t)h / \bar{Y}(t)$$

Estimating $\lambda(t)h$ by $\Delta_h \bar{N}(t) / \bar{Y}(t)$ we get the estimate

$$\widehat{\mathrm{Var}}\left\{\Delta_h \bar{N}(t) / \bar{Y}(t)\right\} = \Delta_h \bar{N}(t) \Big/ \left\{\bar{Y}(t)\right\}^2$$

Adding up across the independent Poisson increments gives the usual variance estimator for the cumulative hazard

$$\widehat{\mathrm{Var}}\left(\hat{\Lambda}(T)\right)=\int_0^T \frac{\mathrm{d}\,\bar{N}(t)}{\left\{\bar{Y}(t)\right\}^2}$$

Now suppose we have a cohort sampling scheme in which patient $i, i=1,2,\ldots,n$ was sampled from a larger cohort with w_i equal to the inverse of the sampling fraction. Accounting for the weighting (and assuming no ties for convenience), the cumulative hazard estimate becomes

$$\hat{\Lambda}^{(w)}(T)=\int_0^T \frac{\sum_{i=1}^n w_i\,\mathrm{d}\,N_i(t)}{\bar{Y}^{(w)}(t)},$$

where

$$\bar{Y}^{(w)}(t)=\sum_{i=1}^n w_i I_{\{t_i\le t\}}$$

In other words, we weight each event count by the sampling weight of the patient that had the event. However, the counts are still generated only from the patients we have sampled, so using the same logic as above we get

$$\mathrm{Var}\left\{\Delta_h\bar{N}(t)\big/\bar{Y}^{(w)}(t)\right\}\approx \bar{Y}(t)\lambda(t)h\Big/\left\{\bar{Y}^{(w)}(t)\right\}^2$$

Estimating $\lambda(t)h$ by $\Delta_h\bar{N}(t)\big/\bar{Y}(t)$ we get the estimate

$$\widehat{\mathrm{Var}}\left\{\Delta_h\bar{N}(t)\big/\bar{Y}^{(w)}(t)\right\}=\Delta_h\bar{N}(t)\Big/\left\{\bar{Y}^{(w)}(t)\right\}^2$$

Our estimate of the cumulative hazard multiplies each of the counts $d\bar{N}(t_i)=\lim_{h\to 0}\Delta_h\bar{N}(t_i)$ by the weight w_i, so the estimated variance of the estimate

$$\hat{\Lambda}^{(w)}(T)=\int_0^T \frac{\sum_{i=1}^n w_i\,\mathrm{d}\,N_i(t)}{\bar{Y}^{(w)}(t)},$$

is

$$\widehat{\mathrm{Var}}\left(\hat{\Lambda}^{(w)}(T)\right)=\int_0^T \frac{\sum_{i=1}^n w_i^2\,\mathrm{d}\,N_i(t)}{\left\{\bar{Y}^{(w)}(t)\right\}^2}$$

////

References

Breslow NE (1972). Contribution to the discussion on the paper by DR Cox, Regression and life tables. *Journal of the Royal Statistical Society, Series B* **34**:216–217.

Cox DR (1972). Regression models and life tables (with discussion). *Journal of the Royal Statistical Society, Series B* **34**:187–220.

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

Therneau TM and Grambsch PM (2000). *Modeling Survival Data*: *Extending the Cox Model*. New York: Springer.

Tsiatis A (1981). A large sample study of the estimates for the integrated hazard function in Cox's regression model for survival data. *Annals of Statistics* **9**:93–108.