



AgEcon SEARCH
RESEARCH IN AGRICULTURAL & APPLIED ECONOMICS

The World's Largest Open Access Agricultural & Applied Economics Digital Library

This document is discoverable and free to researchers across the globe due to the work of AgEcon Search.

Help ensure our sustainability.

Give to AgEcon Search

AgEcon Search

<http://ageconsearch.umn.edu>

aesearch@umn.edu

*Papers downloaded from **AgEcon Search** may be used for non-commercial purposes and personal study only. No other use, including posting to another Internet site, is permitted without permission from the copyright owner (not AgEcon Search), or as allowed under the provisions of Fair Use, U.S. Copyright Act, Title 17 U.S.C.*

Using Aalen's linear hazards model to investigate time-varying effects in the proportional hazards regression model

David W. Hosmer
Department of Biostatistics and Epidemiology
School of Public Health and Health Sciences
University of Massachusetts
715 North Pleasant Street
Amherst, MA 01003-9304 USA
413-545-4532
hosmer@schoolph.umass.edu

Patrick Royston
Cancer Division
MRC Clinical Trials Unit
222 Euston Road
London NW1 2DA
UK
44-20-7670-4736
patrick.royston@ctu.mrc.ac.uk

Abstract. In this paper, we describe a new Stata command, `stlh`, which estimates and tests for the significance of the time-varying regression coefficients in Aalen's linear hazards model; see Aalen (1989). We see two potential uses for this command. One may use it as an alternative to a proportional hazards or other nonlinear hazards regression model analysis to describe the effects of covariates on survival time. A second application is to use the command to supplement a proportional hazards regression model analysis to assist in detecting and then describing the nature of time-varying effects of covariates through plots of the estimated cumulative regression coefficients, with confidence bands, from Aalen's model. We illustrate the use of the command to perform this supplementary analysis with data from a study of residential treatment programs of different durations that are designed to prevent return to drug use.

Keywords: st0024, survival analysis, survival-time regression models, time-to-event analysis

1 Introduction

The Cox proportional hazards model is the most frequently used regression model for the analysis of censored survival-time data, particularly within health sciences disciplines. Stata, in its suite of `st-survival` time programs, has excellent capabilities for fitting the model, as well as options to obtain diagnostic statistics to assess model fit and assumptions. In particular, the vital proportional hazards assumption can be tested using `stphtest` and can be examined graphically using its covariate specific plot option. The problem with the plot is that it is based on the scaled Schoenfeld residuals that are time-point specific and are themselves quite noisy. Even with smoothing, departures from proportionality may be quite hard to determine. It is also not clear how powerful the statistical test in `stphtest` is to detect modest, but, from a subject matter point of view, important departures from proportional hazards. In many applied settings, it may be reasonable to suspect that some covariates may have effects on the hazard

function that are relatively constant effect initially and then fade or end. The converse is also a distinct possibility. The standard procedures and tests have a difficult time diagnosing these situations. We have found plots of the estimated cumulative regression coefficients from a fit of the Aalen linear survival-time model to be a useful adjunct to standard proportional hazards model analyses. The purpose of this paper is to make available a Stata `stlh` command called `stlh` and to illustrate its use.

2 The Aalen linear hazards model

Aalen (1980) proposed a general linear survival-time model, an important feature of which is that its regression coefficients are allowed to vary over time. He discusses issues of estimation, testing, and assessment of model fit in Aalen (1989 and 1993).

2.1 The model

The hazard function at time t for a model containing $p + 1$ covariates, denoted in vector form, $\mathbf{x}' = (1, x_1, x_2, \dots, x_p)$, is

$$h(t, \mathbf{x}, \boldsymbol{\beta}(t)) = \beta_0(t) + \beta_1(t)x_1 + \beta_2(t)x_2 + \dots + \beta_p(t)x_p \quad (1)$$

The coefficients in this model provide the change in hazard at time t , from the baseline hazard function, $\beta_0(t)$, for a one-unit change in the respective covariate, holding all other covariates constant. Note that the model allows the effect of the covariate to change continuously over time. The cumulative hazard function obtained by integrating the hazard function in (1) is

$$\begin{aligned} H(t, \mathbf{x}, \mathbf{B}(t)) &= \int_0^t h(u, \mathbf{x}, \boldsymbol{\beta}(u)) du \\ &= \sum_{k=0}^p x_k \int_0^t \beta_k(u) du \\ &= \sum_{k=0}^p x_k B_k(t) \end{aligned} \quad (2)$$

where $x_0 = 1$ and $B_k(t)$ is called the cumulative regression coefficient for the k th covariate. It follows from (2) that the baseline cumulative hazard function is $B_0(t)$. The model is discussed in some detail in Hosmer and Lemeshow (1999), and the text also includes a review of additional relevant literature. In this paper, the emphasis is placed on using plots of the estimated cumulative regression coefficients to check for possible time-varying covariate effects.

2.2 Estimation

Assume that we have n independent observations of time, a right-censoring indicator variable, assumed to be independent of time conditional on the covariates, and p fixed covariates all denoted by the usual triplet for the i th subject as (t_i, c_i, \mathbf{x}_i) , with $c_i = 0$ for a censored observation and 1 for an event. Aalen's 1989 estimator of the cumulative regression coefficients is a least-squares-like estimator. Denote the data matrix for the subjects at risk at time t_j by an n by $p + 1$ matrix, \mathbf{X}_j , where the i th row contains the data for the i th subject, \mathbf{x}'_i , if the i th subject is in the risk set at time t_j ; otherwise, the i th row is all 0s. Denote by \mathbf{y}_j a n by 1 vector, where the j th element is 1 if the j th subject's observed time, t_j , is a survival time (i.e., $c_j = 1$); otherwise, all the values in the vector are 0. If we consider, in an informal way, the following as an estimator of the vector of the regression coefficient at time t_j ,

$$\hat{\mathbf{b}}(t_j) = (\mathbf{X}'_j \mathbf{X}_j)^{-1} \mathbf{X}'_j \mathbf{y}_j \quad (3)$$

then Aalen's (1989) estimator of the vector of cumulative regression coefficients is

$$\hat{\mathbf{B}}(t) = \sum_{t_j \leq t} \hat{\mathbf{b}}(t_j) \quad (4)$$

Note that the value of the estimator changes only at observed survival times and is constant between observed survival times. Huffer and McKeague (1991) discuss weighted versions of the estimator in (3). The weighted estimator is much more complicated to implement, and it is not clear if it provides better diagnostic power to detect time-varying covariate effects. Thus, it is not used in `stlh`. Also note that the increment in the estimator is computed only when the matrix $(\mathbf{X}'_j \mathbf{X}_j)$ can be inverted; i.e., it is nonsingular. In particular, when there are fewer than $p + 1$ subjects in the risk set, the matrix is singular. Other data configurations can also yield a singular matrix. For example, if the model contains a single dichotomous covariate and all subjects who remain at risk have the same value for the covariate, the matrix will be singular. The `stlh` program checks for this, and estimation stops when $(\mathbf{X}'_j \mathbf{X}_j)$ turns singular.

If we use as an estimator of the variance of $\hat{\mathbf{b}}(t_i)$, the expression

$$\widehat{\text{Var}} \left[\hat{\mathbf{b}}(t_j) \right] = (\mathbf{X}'_j \mathbf{X}_j)^{-1} (\mathbf{X}'_j \mathbf{I}_j \mathbf{X}_j) (\mathbf{X}'_j \mathbf{X}_j)^{-1} \quad (5)$$

then Aalen's (1989) estimator of the covariance matrix of the estimated cumulative regression coefficients at time t may be expressed as

$$\begin{aligned}\widehat{\text{Var}}\left[\widehat{\mathbf{B}}(t)\right] &= \sum_{t_j \leq t} \widehat{\text{Var}}\left[\widehat{\mathbf{b}}(t_j)\right] \\ &= \sum_{t_j \leq t} (\mathbf{X}'_j \mathbf{X}_j)^{-1} (\mathbf{X}'_j \mathbf{I}_j \mathbf{X}_j) (\mathbf{X}'_j \mathbf{X}_j)^{-1}\end{aligned}\quad (6)$$

In equations (5) and (6), the matrix \mathbf{I}_j is an n by n diagonal with \mathbf{y}_j on the main diagonal. It follows from equation (3) and equation (5) that for the k th coefficient,

$$\widehat{\text{Var}}\left[\widehat{b}_k(t_j)\right] = \widehat{b}_k^2(t_j) \quad (7)$$

and

$$\widehat{\text{Var}}\left[\widehat{\mathbf{B}}_k(t)\right] = \sum_{t_j \leq t} \widehat{b}_k^2(t_j) \quad (8)$$

In addition, it follows from (2) and (4) that the estimator of the cumulative hazard function for the i th subject at time t is

$$\widehat{H}(t, \mathbf{x}_i, \widehat{\mathbf{B}}(t)) = \sum_{k=0}^p x_{ik} \widehat{\mathbf{B}}_k(t) \quad (9)$$

and an estimator of the covariate-adjusted survivorship function is

$$\widehat{S}(t, \mathbf{x}_i, \widehat{\mathbf{B}}(t)) = \exp\left[-\widehat{H}(t, \mathbf{x}_i, \widehat{\mathbf{B}}(t))\right] \quad (10)$$

Aalen (1989) notes that it is possible for an estimate of the cumulative hazard in (9) to be negative and to yield a value for (10) that is greater than 1.0. This is most likely to occur for small values of time. One way to avoid this problem is to use zero as the lower bound for the estimator in (9).

The graphical presentation provided to examine for time-varying covariate effects in `stlh` is a plot of $\widehat{\mathbf{B}}_k(t)$ versus t , along with the upper and lower endpoints of a $100(1-\alpha)$ percent pointwise confidence interval,

$$\widehat{\mathbf{B}}_k(t) \pm z_{1-\alpha/2} \times \widehat{\text{SE}}\left[\widehat{\mathbf{B}}_k(t)\right]$$

where $z_{1-\alpha/2}$ is the upper $100(1-\alpha/2)$ percent point of the standard normal distribution, and $\widehat{\text{SE}}\left[\widehat{\text{B}}_k(t)\right]$ is the estimator of the standard error of $\widehat{\text{B}}_k(t)$, obtained as the square root of the variance estimator in (8).

2.3 Testing

Aalen (1989) presents a method for testing the hypotheses that the coefficients in the model are equal to zero. While tests can be made for the overall significance of the model, the `stlh` command implements tests for the significance of individual coefficients. The individual statistics are formed from the components of the vector

$$\widehat{\mathbf{U}} = \sum \mathbf{K}_j \widehat{\mathbf{b}}(t_j) \quad (11)$$

The summation in (11) is over all noncensored times when the matrix $(\mathbf{X}'_j \mathbf{X}_j)$ is nonsingular, and \mathbf{K}_j is a $(p+1) \times (p+1)$ diagonal matrix of weights. Aalen (1989) suggests two choices for weights. One choice mimics the weights used by the Wilcoxon tests and is the number in the risk set at t_j , $\mathbf{K}_j = \text{diag}(m_j)$. His other choice is based on the observation that the estimator in (3) has the same form as the least squares estimator from linear regression. He suggests using weights equal to the square root of the inverse of a least-squares-like variance estimator, namely the inverse of the square root of the diagonal elements of $(\mathbf{X}'_j \mathbf{X}_j)^{-1}$. Lee and Weissfeld (1998) studied the performance of the Aalen's test with these two weights, as well as several others. Based on simulation results, they recommend using weights based on the Kaplan–Meier estimator, $\widehat{S}_{KM}(t)$, at the previous survival time, $\mathbf{K}_j = \text{diag}\left[\widehat{S}_{KM}(t_{j-1})\right]$, with the convention that $\mathbf{K}_1 = \text{diag}\left[\widehat{S}_{KM}(t_0) = 1\right]$ and weights equal to the product of the Kaplan–Meier weights and the Aalen's inverse standard error weights. Lee and Weissfeld (1998) found that these two weight functions were the best at detecting late and early differences, respectively. One obvious choice for weights (that does not seem to have been considered previously) is to mimic the weights for the log-rank test and use $\mathbf{K}_j = \text{diag}(1)$. Based on experience with the log-rank and Wilcoxon tests, we expect that the tests with weights equal to 1 should also be sensitive to later effects, while the test with weights equal to the size of the risk set should be sensitive to early effects.

Since the `stlh` command computes the variance estimator in (6), we use the inverses of the respective diagonal elements of the standard error estimator,

$$\mathbf{K}_j = \text{diag}\left\{\widehat{\text{SE}}\left[\widehat{\mathbf{b}}(t_j)\right]^{-1}\right\} \quad (12)$$

instead of the diagonal elements of $(\mathbf{X}'_j \mathbf{X}_j)^{-1}$.

The variance estimator of \mathbf{U} in (11) is obtained from the variance estimator in (6) and is

$$\begin{aligned}\widehat{\text{Var}}(\widehat{\mathbf{U}}) &= \sum_{t_j} \mathbf{K}_j (\mathbf{X}'_j \mathbf{X}_j)^{-1} (\mathbf{X}'_j \mathbf{I}_j \mathbf{X}_j) (\mathbf{X}'_j \mathbf{X}_j)^{-1} \mathbf{K}_j \\ &= \sum_{t_j} \mathbf{K}_j \widehat{\text{Var}} \left[\widehat{\mathbf{b}}(t_j) \right] \mathbf{K}_j\end{aligned}\quad (13)$$

Tests for significance of individual coefficients use the individual elements of the vector \mathbf{U} and are scaled by the estimator of their standard error obtained as the square root of the appropriate element from the diagonal of the matrix in (13). Aalen remarks that this ratio has approximately the standard normal distribution when the hypothesis of no effect is true and the sample is sufficiently large. Individual standardized test statistics and their significance levels based on the user-requested weights are presented in the output from `stlh`.

In addition to the `stlh` command provided in this paper, we know of two other sources for software to fit and obtain the plots from the Aalen model. Aalen and Fekjær provide macros for use with S-PLUS at <http://www.med.uio.no/imb/stat/addreg> (for the model as described here, as well as an extended model for multivariate survival times, see Aalen et al. (2001)). Klein and Moeschberger (1997) provide macros for use with SAS at <http://www.biostat.mcw.edu/SoftMenu.html>.

3 Syntax

The syntax of `stlh` is similar to other st-survival time regression commands:

```
stlh varlist [if exp] [in range] [, level(#) nograph nomore
      saving(string [, replace] ) testwt(numlist) nodots generate(string)
      tcent(#) graph_options]
```

As with all st commands, one must use `stset` before using `stlh`. The command will accommodate any combination of the following types of data: right censoring, left truncation or delayed entry, and multiple events per subject.

3.1 Required arguments

`varlist` contains the model covariates; the minimum is 1.

3.2 Options

`level(#)` sets `#` to be the confidence level used for pointwise confidence bands in the plots. The default `#` is set by system global macro `$$_level` (default 95, giving 95% confidence bands).

`nograph` suppresses graphical output.

`nomore` suppresses `—more—`, the pausing after displaying each graph.

`saving(string [, replace])` saves the graphs (of cumulative regression coefficients and their confidence bands), with filename(s) `string<xvarname>.gph`. One graph is saved for each covariate in varlist, with `, replace` required to replace existing files with the same names.

`testwt(numlist)` specifies the weights to be used in the tests for no covariate effect. The values in `numlist` may be any or all of the integers 1 through 4, with weights defined as follows:

1. all 1.0
2. size of the risk set
3. Kaplan–Meier estimator
4. Kaplan–Meier estimator divided by the standard error of the regression coefficients

`nodots` suppresses the output of dots displayed for entertainment, with one dot shown per 100 observations processed.

`generate(string)` requests that variables containing the estimated cumulative regression coefficients and their estimated standard error be added to the data. The names of the new variables are `stringB<#>` and `stringS<#>`, where `#` is the order number of the covariate varlist. For example, if `abc` is the chosen string, then `abcB1` contains the estimated cumulative regression coefficients for the first covariate in varlist, and `abcS1` contains the estimated standard errors. The final variable, `abc_cons`, contains results for the constant term (`_cons`).

`tcent(#)` specifies the upper limit on the time axis for the plots. The purpose is to suppress the high variability in the plot expected for long survival times, where the data are sparse and there is little information available to estimate the time-specific coefficients and their variance.

`graph_options` are any of the options allowed with `graph`, `twoway`, except for `sort`, `pen()`, `symbol()`, `connect()`, `ttitle()`, `yline()`, and `saving()`. To customize graphs, one needs to save the relevant quantities (see the `generate()` option) and then recreate the confidence bands as cumulative regression coefficient $\pm z$ * standard error, where $z = 1.96$ for the default 95% confidence band.

In cases where there are ties in the survival times, the results can depend on the order of the tied survival. In order to have invariant results, in this case, `stlh` sorts the data in a unique and reproducible fashion. This involves first sorting the names of the covariates lexicographically using the method described by Royston (2001), and then including the covariates in the specified order in a sort of the failure times.

4 An example

To illustrate the use of the Aalen linear survival-time model as implemented in `stlh`, we consider a subset of the data and variables from the UIS Study described in Hosmer and Lemeshow (1999, Section 1.3). Briefly, this study is a randomized trial of residential treatment programs of two different lengths or durations for drug abuse. The time variable records the number of days from randomization to treatment until self-reported return to drug use, lost to follow-up, or end of the study. The right censoring variable is equal to 1 for return to drug use and 0 otherwise.

4.1 Proportional hazards analysis

As an example, we fit the proportional hazards model containing age, the subject's Beck depression score at randomization, and an indicator variable for treatment (0 = shorter duration, 1 = longer duration). The data for the 575 subjects used in this example are in the Stata data file `uisaalen.dta`. There are a total of 628 subjects in the main dataset and 12 variables. A text file containing all the data, as well as text files for all the other datasets used in Hosmer and Lemeshow (1999), can be downloaded from the John Wiley & Sons Inc. ftp site: ftp://ftp.wiley.com/public/sci_tech_med/survival, or from the survival section of the dataset archive at the University of Massachusetts, <http://www-unix.oit.umass.edu/~statdata>.

Before fitting the models, we centered age at 32.4 years and Beck score at 17.4, their respective means. Thus, the baseline cumulative hazard is for a subject of average age, average Beck score, and randomized to the shorter intervention.

The results of fitting the proportional hazards model are shown below. Note that we saved the Schoenfeld residuals in order to use `stphtest`. Following the fit of the model are the results of `stphtest` performed on the log-time and time scale. In addition, we present results from a fit including continuous time-varying interactions between each covariate and time by utilizing the `tv` and `te` options of `stcox`.

```
. stset time, failure(censor)
. stcox age_c beck_c treat, nolog nohr sch(sch*) sca(sca*)
      failure _d:  censor
      analysis time _t:  time
                   id:  id
```

```

Cox regression -- Breslow method for ties
No. of subjects =      575          Number of obs   =      575
No. of failures =      464
Time at risk    =     138900
Log likelihood   =    -2657.017
LR chi2(3)      =      13.94
Prob > chi2     =      0.0030

```

_t _d	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
age_c	-.0130252	.007501	-1.74	0.082	-.0277268	.0016764
beck_c	.00989	.0048615	2.03	0.042	.0003617	.0194184
treat	-.2394376	.0930941	-2.57	0.010	-.4218986	-.0569765

```
. stphtest, detail log
```

```
Test of proportional hazards assumption
```

```
Time: Log(t)
```

	rho	chi2	df	Prob>chi2
age_c	0.02248	0.22	1	0.6380
beck_c	-0.06353	1.74	1	0.1870
treat	0.06430	1.90	1	0.1684
global test		3.90	3	0.2724

```
. stphtest,detail
```

```
Test of proportional hazards assumption
```

```
Time: Time
```

	rho	chi2	df	Prob>chi2
age_c	0.03007	0.40	1	0.5291
beck_c	-0.06057	1.58	1	0.2084
treat	0.11437	6.00	1	0.0143
global test		8.00	3	0.0459

The coefficients for Beck score and treatment are significant with $p < 0.05$, while age is significant at the 10 percent level. The tests for proportional hazards using `stphtest` on the log-time scale are not significant overall or for each of the three covariates. When we test for proportional hazards on the time scale, we see that the test is significant overall, $p = 0.045$, largely due to the significance of the test for treatment where $p = 0.0143$. The plots of the scaled Schoenfeld residuals and their smooth on the time scale are shown in Figure 1, Figure 2, and Figure 3.

```
. stphtest, plot(age_c)
```

(Continued on next page)

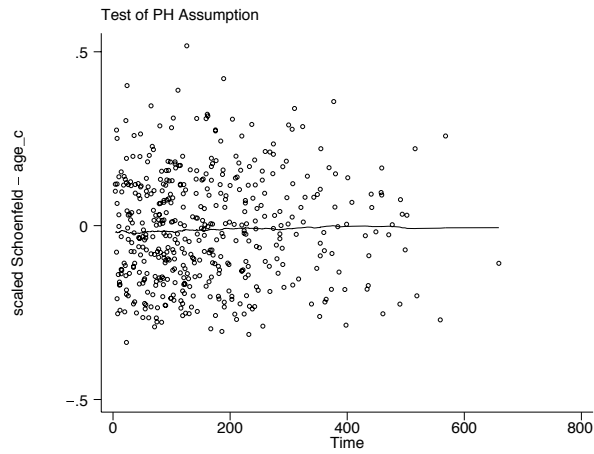


Figure 1: Plot of the raw and smoothed scaled Schoenfeld residuals for centered age.

```
. stphtest, plot(beck_c)
```

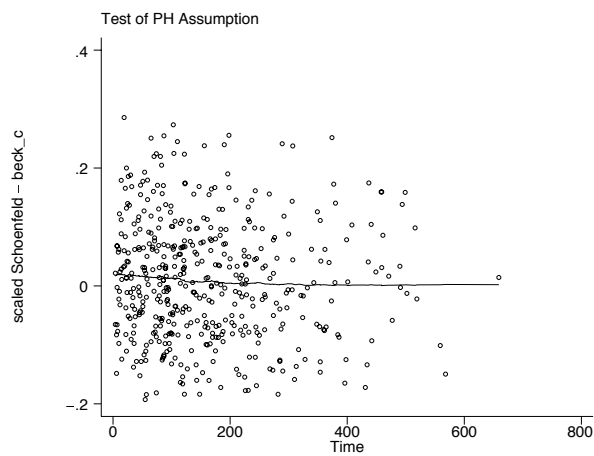


Figure 2: Plot of the raw and smoothed scaled Schoenfeld residuals for centered Beck score.

```
. stphtest, plot(treat)
```

(Continued on next page)

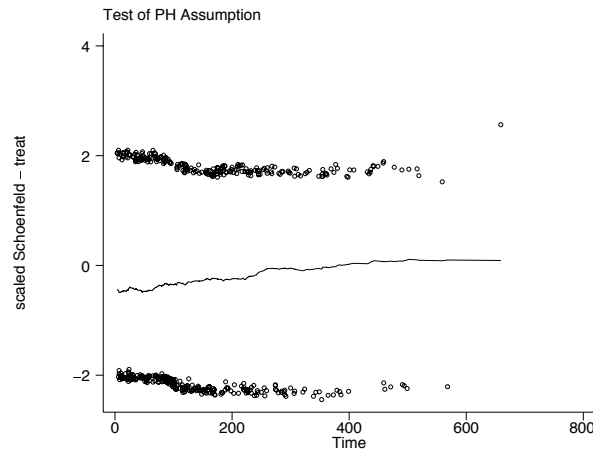


Figure 3: Plot of the raw and smoothed scaled Schoenfeld residuals for treatment.

The smoothed line in the plot for age in Figure 1 has a slope approximately equal to zero, suggesting that there may be no time-varying effect, and this is in agreement with the test. The plot for Beck score in Figure 2 appears to have a slight negative slope, suggesting the potential for time-varying effect. The smoothed line in the plot for treatment in Figure 3 has a definite positive slope, suggesting that treatment has a diminishing time-varying effect.

Winnett and Sasieni (2001) show that the scaling of the Schoenfeld residuals used by Stata in the plots from `stphtest` may yield plots that do not demonstrate the correct nonproportional effect when one is present. Their work shows that this is most likely to occur when the range of a covariate changes over the risk sets. In this case, the covariance matrix of the Schoenfeld residuals is not approximately constant and thus is not well approximated by its average. Plots, not shown, of the age and Beck score versus time show that the range is approximately constant over time; thus, the smooth in Figure 1 and Figure 2 should provide a good estimate of any nonproportional effects. Since the treatment covariate is dichotomous, the range is always 1.0 when estimation is possible. Winnett and Sasieni note that the computational burden of using the more sensitive scaling suggested in their paper is not severe and thus could be easily incorporated into a future version of `stphtest`.

To complete the analysis for nonproportional hazards, we display the fit of a model that adds interactions between model covariates and analysis time. We note that the p -values for the Wald tests for the interaction variables are similar to the p -values from `stphtest`. The results of a similar analysis on the log time scale agree with the ones presented for `stphtest` and are not shown here.

```
. stcox age_c beck_c treat, nolog nohr tvc( age_c beck_c treat) texp(_t)
failure_d: censor
      analysis time _t: time
              id: id

Cox regression -- Breslow method for ties
No. of subjects =          575                Number of obs =          575
No. of failures =          464
Time at risk   =         138900
Log likelihood =   -2653.0912                LR chi2(6) =          21.79
                                                Prob > chi2 =          0.0013
```

	_t		Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
	_d						
rh							
	age_c		-.0189407	.0125671	-1.51	0.132	-.0435718 .0056904
	beck_c		.0178941	.0081022	2.21	0.027	.002014 .0337741
	treat		-.5467554	.1576506	-3.47	0.000	-.855745 -.2377658
t							
	age_c		.0000311	.0000603	0.52	0.606	-.0000871 .0001492
	beck_c		-.0000483	.0000402	-1.20	0.230	-.0001271 .0000306
	treat		.0018746	.0007857	2.39	0.017	.0003347 .0034145

note: second equation contains variables that continuously vary with respect to time; variables are interacted with current values of _t.

Thus, at the conclusion of what one might call the standard analysis to examine for proportional hazards, we have significant evidence of nonproportional hazards for treatment, some graphical evidence for Beck score that is not supported by `stphtest`, and no evidence of a nonproportional hazard in age.

The problem we face now is to try and figure out from the plots the nature of the time dependency in the effect of treatment and possibly Beck score. The problem is that the smoothed line in the plots gives few clues as to how to parameterize the time-varying effect. As we demonstrate, plots of the estimated cumulative regression coefficients from Aalen's linear survival-time model can be quite helpful in identifying the form of time-varying effects in the proportional hazards model.

4.2 Aalen linear hazards model analysis: estimation

Before looking at the plots for the example, we discuss their expected behavior. Suppose that the time-varying coefficient for a covariate in Aalen's model in equation (1) is constant, $b(t) = \alpha$. It follows that the cumulative coefficient at time t is $B(t) = \alpha t$. In this case, the plot of the cumulative regression coefficient versus time is a straight line with slope α . Now suppose that the covariate has no further effect on the hazard after, say, 200 days. The plot after 200 days will be constant and equal to 200α . If the effect, α , is significant, then the confidence bands are expected not to include zero. Similar arguments can be used to describe a late effect α with no early effect, as well as different effects in different time intervals.

After Aalen first introduced his model, there was a flurry of additional research on its properties. Of particular relevance to this paper is the work by Henderson and Milner (1991). They demonstrated that even under proportional hazards, a covariate exhibits a slight curve, nonlinearity, in the Aalen plots. They show that the more significant the effect in the proportional hazards model, the more curved the plot. While we must keep this in mind when examining the Aalen plots, we do so with the knowledge that tests and plots based on the proportional hazards model have already suggested that some covariates may have nonproportional hazards. More importantly, any time-varying effect must make contextual (in this case, clinical) sense.

The plots of the cumulative regression coefficients that result from using `stlh` to fit the three-covariate model are shown below. For the age and Beck score, we present the plots from `stlh`. The time-varying effect for treatment is a bit more complex. In order to better focus the discussion, we present an annotated plot obtained by using `graph` with generated variables. In these data, the maximum observed survival time when the matrix $(\mathbf{X}'_j \mathbf{X}_j)$ was nonsingular was 569 days. However, we restrict the plots to the interval 0 to 377 days, the 75th percentile of the survival time. It is our experience that the plot is highly variable beyond this point due to few subjects still at risk.

The plot of the cumulative coefficient for age in Figure 4 decreases linearly and flattens a bit after about 150 to 180 days. Note that the upper confidence band crosses back and forth across the zero line, suggesting that age might not be significant in the Aalen model. We discuss tests for covariate effects in the Aalen model after the plots.

```
. stlh age_c beck_c treat, xlabel(0,90,180,270,377) ltitle("Hazard") /*
*/ testwt(1 2 3 4) b1title(" ") b2title("Time") gen(uis)
```

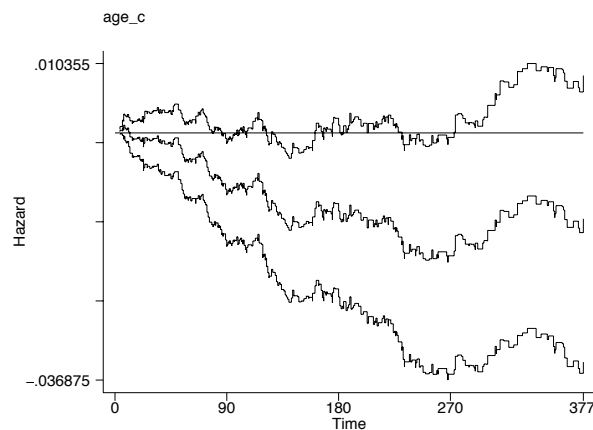


Figure 4: Plot of the estimated cumulative regression coefficient for centered age and for the pointwise 95 percent confidence bands.

The plot of the cumulative regression coefficient for Beck score in Figure 5 increases in a curvilinear manner for the first 180 or so days and then has roughly zero slope.

This plot suggests that Beck score may have an early effect, up to 180 days, and no late effect. Note that the lower confidence band does not include the zero line for most of the first 180 days and does so after 180 days. This is consistent with the pattern of an early, but no late effect for Beck score, or in terms of the time-varying coefficient

$$b_{beck}(t) = \begin{cases} \alpha & \text{if } t \leq 180 \\ 0 & \text{if } t > 180 \end{cases} \quad (14)$$

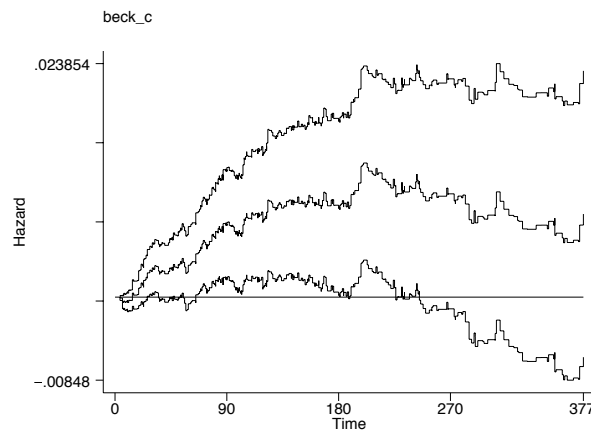


Figure 5: Plot of the estimated cumulative regression coefficient for centered Beck score and the pointwise 95 percent confidence bands.

The pattern in the plot of the cumulative regression coefficient for treatment in Figure 6 is much more complex. Recall that the shorter treatment was for 90 days, and the longer for 180 days planned duration. For the first 75 days, the slope is effectively zero, and the upper confidence band lies above the zero line. From days 75 to 90, the slope is negative, but the upper confidence band still lies above the zero line. The curve continues with the same negative slope until about 180 days, after which it is again approximately zero. This pattern suggests that the time-varying coefficient in the Aalen model is zero up to 90 days, indicating no early effect; is nonzero and constant from 90 to 180 days, indicating a middle effect; and is zero after 180 days, indicating no late effect. Specifically,

$$b_{treat}(t) = \begin{cases} 0 & \text{if } t \leq 90 \\ \alpha & \text{if } 90 < t \leq 180 \\ 0 & \text{if } t > 180 \end{cases} \quad (15)$$

These observations make sense from a clinical point of view, as they agree with the two different durations of planned treatment. It is not surprising that while subjects in both planned durations are under treatment, there would be no effect for the longer

treatment. (The actual study is more complicated than this, but this discussion is accurate enough for the purposes of this paper.) From 90 to 180 days, only subjects in the longer treatment could still be in the residential program. Thus, we expect that there may be an effect in this time interval. After 180 days, none of the subjects remain on active treatment. We might expect some continuing benefit of the longer treatment, but there is none. Refitting the Aalen model for follow-up times greater than 180 days confirms this observation.

```
. rename uisB3 Btreat
. gen Btreat_l= Btreat-1.96* uisS3
. gen Btreat_u= Btreat+1.96* uisS3
. graph Btreat_l Btreat_u Btreat time if time <=377, c(111) s(iii)/*
  / xline(60,75,90,180) xlabel(0,60,75,90,180,270,377)/*
  /* yline(0) ylabel(-0.68,0,0.061) b1title(" ") b2title("Time")/*
  /* l1title("Hazard") t1title(" ")
```



Figure 6: Plot of the estimated cumulative regression coefficient for treatment and the pointwise 95 percent confidence bands.

4.3 Proportional hazards analysis with time-varying covariates

The next step in the proportional hazards analysis is to create the time-varying covariates for Beck score and treatment as suggested by the Aalen model plots and the hypothesized time-varying coefficients in equation (14) and equation (15). For Beck score, we must create two new time-varying covariates:

$$beck_{early}(t) = \begin{cases} 1 & \text{if } t \leq 180 \\ 0 & \text{if } t > 180 \end{cases}$$

and

$$beck_{late}(t) = \begin{cases} 1 & \text{if } t > 180 \\ 0 & \text{if } t \leq 180 \end{cases}$$

For treatment, we must create three new time-varying covariates:

$$treat_{early}(t) = \begin{cases} 1 & \text{if } t \leq 90 \\ 0 & \text{if } t > 90 \end{cases}$$

$$treat_{mid}(t) = \begin{cases} 1 & \text{if } 90 \leq t \leq 180 \\ 0 & \text{if } t < 90 \text{ or } t > 180 \end{cases}$$

and

$$treat_{late}(t) = \begin{cases} 1 & \text{if } t > 180 \\ 0 & \text{if } t \leq 180 \end{cases}$$

We create these covariates using the `stsplit` command to split the records at 90 and 180 days. The newly created variables, `splt1` and `splt2`, are then used to create the time-varying covariates. The Stata commands are as follows:

```
. stsplit splt1, at(90)
(414 observations (episodes) created)
. stsplit splt2, at(180)
(277 observations (episodes) created)
. replace splt1 = 1 if splt1 > 0
(691 real changes made)
. replace splt2 = 1 if splt2 > 0
(277 real changes made)
. gen bck_early=beck_c*(1-splt2)
. gen bck_late=beck_c*splt2
. gen trt_early=treat*(1-splt1)
. gen trt_mid=treat*splt1*(1-splt2)
. gen trt_late=treat*splt2
```

The next step is to fit the model after replacing `beck_c` and then `treat` by their time-varying versions. We save the Schoenfeld residuals in order to use `stphtest` to test for proportional hazards.

(Continued on next page)

```
. stcox age_c bck_early bck_late trt_early trt_mid trt_late, nohr nolog /*
*/ noshow sch(sch*)sca(sca*)
Cox regression -- Breslow method for ties
No. of subjects =          575                Number of obs   =       1266
No. of failures =          464
Time at risk   =          138900
Log likelihood = -2653.1897                LR chi2(6)        =       21.59
                                                Prob > chi2       =       0.0014
```

_t _d	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
age_c	-.0136093	.0075037	-1.81	0.070	-.0283162	.0010976
bck_early	.014278	.0060733	2.35	0.019	.0023745	.0261814
bck_late	.0027401	.0081245	0.34	0.736	-.0131836	.0186637
trt_early	-.2725956	.1588572	-1.72	0.086	-.58395	.0387588
trt_mid	-.5427174	.1730745	-3.14	0.002	-.881937	-.2034977
trt_late	.0368758	.157425	0.23	0.815	-.2716715	.3454232

```
. stphtest, detail
```

```
Test of proportional hazards assumption
```

```
Time: Time
```

	rho	chi2	df	Prob>chi2
age_c	0.02533	0.28	1	0.5955
bck_early	-0.02399	0.25	1	0.6201
bck_late	0.00931	0.04	1	0.8455
trt_early	-0.00310	0.00	1	0.9467
trt_mid	0.02973	0.40	1	0.5266
trt_late	0.05033	1.18	1	0.2768
global test		2.16	6	0.9040

The results of the fit support the observation of an early, but no late effect for Beck score, as the p -values are 0.019 for `bck_early` and 0.736 for `bck_late`. The results do not completely support our observations on the time-varying effect of treatment. The p -values for the three coefficients are 0.086, 0.002, and 0.815. The interpretation is that there is an indication of some possible early effect, at the 10 percent level, and a highly significant treatment effect between 90 and 180 days. There is no significant late effect. If we define the early and mid time-varying covariates for treatment using 75 days as the cut-point, the three p -values are (output not presented) 0.293, <0.001, and 0.815, respectively. Both analyses, 75 and 90 day cut-points, support a significant treatment effect in the time interval when subjects in the longer duration treatment group are in the residential treatment facility, and the ones on the shorter duration treatment are not.

The results from `stphtest` support proportional hazards of all model covariates on both the time and the log-time scale (output not presented).

The next step in the proportional hazards model analysis would be to use the results from the fitted model to estimate time-interval-specific hazard ratios for Beck score and

treatment and an overall hazard ratio for age. We do not show these results, as they are standard steps that we assume the reader is quite comfortable in performing.

In summary, the plots of the estimated cumulative regression coefficients with confidence bands from the Aalen linear survival-time model have proven to be a useful adjunct to the standard analysis for proportional hazards. We emphasize “adjunct”, as we know from results of Henderson and Milner (1991) that nonlinearity in the Aalen model plots can occur for covariates that have proportional hazards. Thus, it is vital that any derived time-varying covariates have a sound grounding in the science of the problem being studied.

4.4 Aalen linear hazards model analysis: testing

The last point we touch on in this note is tests for no covariate effect in the Aalen model, $H_0 : b_k(t) = 0$ for $k = 0, 1, 2, K, p$. As we noted, the `stlh` command supports four weight functions: (1) weights equal to 1, (2) weights equal to the size of the risk set, (3) weights equal to the Kaplan–Meier estimator at the previous survival time, and (4) weights equal to the product of the third weight and the inverse of the standard deviation of the time-specific Aalen model coefficient. The results from the test portion of the `stlh` command and each of the four weight functions follow.

```
. stlh age_c beck_c treat, test(1 2 3 4) nograph
```

```
Graphs and tests for Aalen's Additive Model
```

```
-----
Model:  age_c beck_c treat
Obs:    1266
```

```
Test 1: Uses Weights Equal to
1.0
```

Variable	z	P
age_c	-1.323	0.186
beck_c	1.385	0.166
treat	0.551	0.582
_cons	12.515	0.000

```
Test 2: Uses Weights Equal to
the Size of the Risk Set
```

Variable	z	P
age_c	-2.288	0.022
beck_c	2.515	0.012
treat	-2.902	0.004
_cons	14.959	0.000

```
Test 3: Uses Weights Equal to
Kaplan-Meier Estimator at Time t-
```

Variable	z	P
age_c	-1.932	0.053
beck_c	2.167	0.030
treat	-0.673	0.501
_cons	15.301	0.000

Test 4: Uses Weights Equal to
(Kaplan-Meier Estimator at Time t-)/(Std. Dev of the Time-varying Coefficient)

Variable	z	P
age_c	-2.001	0.045
beck_c	1.201	0.230
treat	-1.696	0.090
_cons	12.242	0.000

When we test using weights equal to 1, none of the tests for covariate effect are significant. The test appears to be picking up the fact that there is either no effect or no additional effect after 180 days.

When we test using weights equal to the size of the risk set, all of the tests for covariate effect are significant. Here the tests seem to pick up the fact that all covariates have some effect in the interval from 0 to 180 days.

The test results obtained when using the Kaplan–Meier weights or weights equal to the product of the Kaplan–Meier weights and the inverse of the estimated standard deviation of $\hat{b}(t)$ yield results that are contrary to the observations of Lee and Weissfeld (1998). The test using Kaplan–Meier weights detects the early difference in Beck score, but not the middle effect in treatment. The reverse is true when using weights equal to the product of the Kaplan–Meier weights and the inverse of the estimated standard deviation of $\hat{b}(t)$. We are not sure why this is the case. It certainly warrants further simulation studies, as Lee and Weissfeld (1998) only considered models containing a single dichotomous covariate.

5 Summary

In this paper, we have shown how a new Stata command, `stlh`, can be used as an adjunct to the traditional proportional hazards analysis to help identify the nature of time-varying effects of covariates. The example shows that this analysis can be particularly useful when covariates have constant but different effects in different time intervals. Since the basic plot of the estimated cumulative regression coefficients can display curvature when covariate effects are proportional, considerable care must be taken when interpreting their shape. Any identified time-varying effect must have a sound contextual basis before being added to the proportional hazards model.

6 Acknowledgment

The authors would like to thank Peter Sasieni for helpful discussions that improved the paper and for referring us to related work by Angela Winnett and himself.

(Continued on next page)

7 References

- Aalen, O. O. 1980. A model for non-parametric regression analysis of counting processes. *Lecture Notes in Statistics* 2: 1–25.
- . 1989. A linear regression model for the analysis of life times. *Statistics in Medicine* 8: 907–925.
- . 1993. Further results on the non-parametric linear regression model in survival analysis. *Statistics in Medicine* 13: 1569–1588.
- Aalen, O. O., O. Borgan, and H. Fekjær. 2001. Covariate adjustment of event histories estimated from Markov chains: The additive approach. *Biometrics* 57: 993–1001.
- Henderson, R. and A. Milner. 1991. Aalen plots under proportional hazards. *Applied Statistics* 40: 401–409.
- Hosmer, D. W., Jr. and S. Lemeshow. 1999. *Applied Survival Analysis: Regression Modeling of Time to Event Data*. New York: John Wiley and Sons.
- Huffer, F. and I. W. McKeague. 1991. Weighted least squares estimation of Aalen's additive risk model. *Journal of the American Statistical Association* 86: 114–129.
- Klein, J. P. and M. L. Moeschberger. 1997. *Survival Analysis: Techniques for Censored and Truncated Data*. New York: Springer-Verlag.
- Lee, E. and L. A. Weissfeld. 1998. Assessment of covariate effects in Aalen's additive hazard model. *Statistics in Medicine* 17: 983–998.
- Royston, P. 2001. Sort a list of items. *Stata Journal* 1(1): 101–104.
- Winnett, A. and P. Sasieni. 2001. A note on scaled Schoenfeld residuals for the proportional hazards model. *Biometrika* 88: 565–571.

About the Authors

David Hosmer is an emeritus professor of biostatistics in the Department of Biostatistics and Epidemiology of the University of Massachusetts School of Public Health and Health Sciences in Amherst, Massachusetts.

Patrick Royston is a medical statistician of 25 years of experience, with a strong interest in biostatistical methodology and in statistical computing and algorithms. At present he works in clinical trials and related research issues in cancer. Currently he is focusing on problems of model building and validation with survival data, including prognostic factors studies, on parametric modeling of survival data and on novel trial designs.