A SAS Macro For Estimation Of Direct Adjusted Survival Curves Based On A Stratified Cox Regression Model

Xu Zhang

Department of Mathematics and Statistics, Georgia State University, Atlanta, GA 30302, U.S.A.

Fausto R. Loberiza

Department of Internal Medicine, University of Nebraska, Omaha, NE 68198, U.S.A

John P. Klein and Mei-Jie Zhang

1 Introduction

The Cox [1] model or log rank test [2] is commonly used in medical studies to compare the survival of patients on di erent treatments. In randomized clinical trials comparisons between treatments are direct and summary survival curves produced by using a Kaplan-Meier [3] technique are used to represent the survival experience of a patient given a specific treatment. These unadjusted curves represent the typical patient since randomization assures where $\hat{\Lambda}_{0i}(t)$ is the estimated cumulative hazard function, can be computed in SAS using the BASELINE command. This curve represents the survival experience of a patient with a prognostic index $\boldsymbol{\beta}^T\overline{\boldsymbol{Z}}$ equal to the average prognostic index of all patients. This method, while easy to implement in practice, has several drawbacks. First, the covariate value for the average patient may be quite meaningless for categorical variables. For example, if one of the covariates is gender coded as 0 for male and 1 for female, the meaning of patient with a sex covariate of 0.4 is hard to interpret. Second, as discussed in Thomsen et al. [6] this method does not account for the sample variability in the prognostic indicator from individual to individual.

The second method often called the "direct adjusted survival curve" by among others Chang et al. [7], Makuch [8] and Gail and Byar [9] averages the estimated survival curves for each patient. That is

$$\hat{S}_i(t) = \frac{1}{n}\sum_{l=1}^{n} \exp\left[-\hat{\Lambda}_{0i}(t)e^{\hat{\boldsymbol{\beta}}^T \boldsymbol{Z}_l}\right]. \tag{3}$$

This method averages survival curves for each patient in the sample rather than the covariates and produces a more representative survival curve. Lee et al. [10] and Ghali et al. [11] provided programs in SAS, STATA, and S-plus for deriving such curves.

Other methods have been proposed to make adjustments when there is a treatment group present. Nieto and Coresh [12] provided a comparison of these methods. Cole and Hernán [13] discussed a technique that uses weights from a logistic regression of the covariates on treatment indicator to make adjustments to the survival estimator. They provided a SAS macro to implement this method. The macro seems to require only two treatment groups and does not provide estimates of the precision of the survival estimates.

Existing programs to compute the direct adjusted survival curves are of limited utility since they do not provide estimates of the uncertainty in the estimators such as the standard errors or confidence intervals. These programs have focused on the case where the treatment hazards ratio, after adjustment for covariates, are constant. They are not applicable when

We have implemented a SAS macro that computes the directed adjusted survival function for treatment groups based on either an unstratified Cox model or a stratified Cox model. The macro also produces standard errors of the estimates of survival and standard errors of the di erence in survival between pairs of treatment groups. The standard errors of the di erence can be used to make pointwise comparisons of treatment groups.

In Section 2 we review the estimation techniques for the direct adjusted survival estimators and their standard errors based on a stratified Cox model. The variance estimations for the direct adjusted survival probabilities and the di erences of the direct adjusted survival probabilities are given in the appendix. We describe the SAS macro and its output in Section 3. In Section 4 we utilize the SAS macro to analyze stem cell transplant data and Ewing's sarcoma data. Discussions are given in Section 5.

2 Estimating the Direct Adjusted Survival Curve

Let the observations on subject j of treatment group i be $\{T_{ij}$

for cloglog transformation and arcsine-square root transformation can be found in Klein and Moeschberger [2].

In some studies a comparison of the direct adjusted survival probabilities of two treatments at a given time is of interest [9]. For example, the 5-year survival rate is an important parameter used to evaluate different treatments in a cancer study. Pointwise comparison of treatment probabilities can be implemented by constructing a confidence interval for the difference of the two adjusted survival probabilities [9]. In the Appendix we give the variance estimator for $\hat{S}_i(t) - \hat{S}_j(t)$, which is denoted by $\hat{\sigma}^2_{i,j}(t)$. Then a $(1-\alpha)100\%$ confidence interval for $S_i(t) - S_j(t)$ is given by

$$\left(\hat{S}_i(t) - \hat{S}_j(t)\right) \pm z_{\alpha/2}\hat{\sigma}_{i,j}(t).$$

3 The SAS Macro

We have written a SAS macro to compute the direct adjusted survival curves. The macro reports $\hat{S}_i(t), \hat{\sigma}_{i,i}(t)$ for $i = 1, \cdots, K$ and $\hat{\sigma}_{i,j}(t)$ for $1 \le i < j \le K$.

The macro requires a SAS data set with the following variables: 1) a variable with the failure time; 2) an indicator variable that indicates if an event has occurred (coded as 1 for an event and 0 for censoring); 3) a variable that indexes the treatments (coded as $1, \ldots, K$); and 4) variables for all risk factors.

Suppose the macro is saved as a SAS file with the filename ADJSURV.sas. One can save a copy of the file in the current working directory, and then use the following SAS statement to load the macro into the current program.

```
%INCLUDE 'ADJSURV.sas';
```

The macro will be invoked by running the following statement.

```
%ADJSURV(inputdata, time, event, group, covlist, model, outdata);
```

2 for an unstratified Cox model;

`outdata` the SAS output data set name;

The results of the macro are saved in the SAS output data set "`outdata`" and printed in

%INCLUDE 'ADJSURV.sas';

%ADJSURV(transplant,time,event,group, stage chemo1 chemo2 LDH1 LDH2 kscore DX2T1 DX2T2 age year1 year2, 1, out);

This provided an output data set with the adjusted survival estimates based on the stratified Cox model. To obtain the estimates for the unstratified Cox model we replaced "1" by "2" in the second to the last argument of the macro. Both the unstratified Cox model and the stratified Cox model estimates took approximately 28 seconds of CPU time on a SUN BLADE 2500 workstation with 1.28 GHZ processor and 2GB of RAM. The unadjusted Kaplan-Meier curves are shown in Figure 1. Figure 3 and 5 show the direct adjusted survival curves based on the unstratified and stratified Cox model, respectively. The stratified Cox model is more appropriate since the crude survival curves in Figure 1 show nonproportional hazards between transplant groups. For the purpose of comparison, the adjusted survival curves (2) for a patient with mean values of the covariates are also provided (Figure 2 and 4). It should be noted that this is a hypothetical patient and it is di cult to interpret the survival probabilities for such a patient.

[Insert Figure 1-5 here]

For both models the data set "out" was produced. This data set included 10 variables. Part of the output based on the stratified Cox model is list below:

Obs	time	surv1	se1	surv2	se2	surv3	se3	se12	se13	se23
					......					
265	55.2	0.565	0.022	0.691	0.041	0.520	0.041	0.046	0.047	0.059
266	57.2	0.561	0.023	0.691	0.041	0.520	0.041	0.046	0.047	0.059
267	59.1	0.561	0.023	0.677	0.043	0.520	0.041	0.048	0.047	0.060
268	59.6	0.556	0.023	0.677	0.043	0.520	0.041	0.048	0.047	0.060
269	60.0	0.551	0.023	0.677	0.043	0.520	0.041	0.048	0.047	0.060
					......					

From this output we can obtain the 5 year (60 month) survival for the three treatments. These are shown in Table 1 for both models, for the unadjusted Kaplan-Meier estimator and for the adjusted survival curve (2) available by using the baseline statement in PROC PHREG.

[Insert Table 1 here]

5 Discussions

We have presented a SAS macro to compute the direct adjusted survival curves and the variances based on a stratified or unstratified Cox model. These curves, as discussed in Thomsen et al. [6], provide more realistic estimates of the "average" survival probability for a treatment by presenting the average survival curve if each patient in the sample had received a given treatment. The average survival curve presented here is equal to $E_{Z}\{\hat{S}(t|Z)\}$, where the expectation is taken over the empirical distribution of Z based on the complete sample. In practice one could take this expectation over different distribution for Z to obtain the average survival function for a hypothetical population with this distribution of the covariates.

The macro provides estimates of the standard errors of the difference in survival curves at all time points. These can be used to make comparisons on survival between different treatment arms at a fixed time point. One could use these estimates to construct pointwise confidence intervals for the difference in adjusted survival but we strongly recommend a confidence band that accounts for multiple testing in looking at a range of time points. The technique for constructing a confidence band for the difference in adjusted survival can be found in Zhang and Klein [14].

The macro can be found on our website at
"`http://www.biostat.mcw.edu/software/SoftMenu.html`".

Acknowledgments

Dr. Klein and Dr. Zhang's research was supported by National Cancer Institute grant 2 R01 CA54706-10. We thank the CIBMTR for providing us the data.

References

[1] Cox, D.R. (1972). Regression models and life-tables (with discussion). *Journal of the Royal Statistical Society: Series B* 34, 187–220.

[2] Klein, J.P. and Moeschberger, M.L. (2003). *Survival Analysis: Techniques for Censored and Truncated Data*, 2nd Edition. New York: Springer-Verlag.

[3] Kaplan, E.L. and Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association* 53, 457–481.

[4] Breslow, N.E. (1975). Analysis of survival data under the proportional hazards model. *International Statistical Review* 43, 45–58.

[5] Neuberger, J., Altman, D.G., Christensen, E., Tygstrup, N. and Williams, R. (1986). Use of a prognostic index in evaluation of liver transplantation for primary biliary cirrhosis. *Transplantation* 41, 713–716.

[6] Thomsen, B.L., Keiding, N. and Altman, D.G. (1991). A note on the calculation of expected survival, illustrated by the survival of liver transplant patients. *Statistics in Medicine* 10 733–738.

[7] Chang, I.M., Gelman, R. and Pagano, M. (1982). Corrected group prognostic curves and summary statistics. *Journal of Chronic Diseases* 35, 669–674.

[8] Makuch, R.W. (1982). Adjusted survival curve estimation using covariates. *Journal of Chronic Diseases* 35, 437–443.

[9] Gail, M.H. and Byar, D.P. (1986). Variance calculations for direct adjusted survival curves, with applications to testing for no treatment e ect. *Biometrical Journal* 28, 587–599.

[10] Lee, J., Yoshizawa, C., Wilkens, L. and Lee, H.P. (1992). Covariance adjustment of survival curves based on Cox's proportional hazards regression model. *Computer Applications in the Biosciences* 8, 23–27.

[11] Ghali, W.A., Quan, H., Brant, R., et al. (2001). Comparison of 2 methods for calculating adjusted survival curves from proportional hazards models. *Journal of the American Medical Association*

Similarly, the variance of □

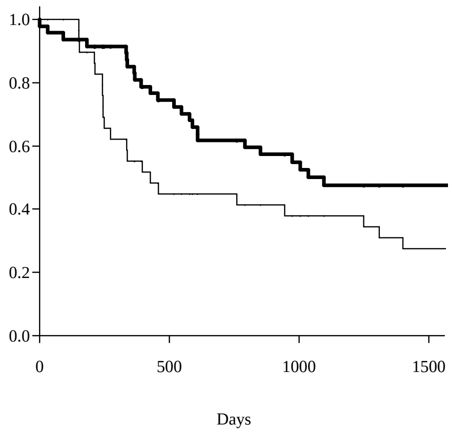
1.0
0.8
0.6
0.4
0.2
0.0
0
500
1000
1500
Days

Table 1. Estimate of survival and standard error at 5 years post transplant.

Type of Transplant	Kaplan-Meier	Direct Adjusted		Via Baseline Command	
		Stratified	Unstratified	Stratified	Unstratified
Unpurged Autologous	0.55 (0.03)	0.55 (0.02)	0.57 (0.02)	0.57 (0.03)	0.59 (0.02)
Purged Autologous	0.62 (0.05)	0.68 (0.04)	0.67 (0.04)	0.70 (0.05)	0.70 (0.04)
Allogeneic	0.52 (0.04)	0.52 (0.04)	0.42 (0.04)	0.53 (0.05)	0.41 (0.05)

Table 2. Estimated di erence in survival at 5 years post transplant.

Di erence	Kaplan-Meier		Stratified		Unstratified	
	Estimate (95% CI)	P	Estimate (95% CI)	P	Estimate (95% CI)	P
Unpurged − Purged	-0.07 (-0.18, 0.03)	0.18	-0.13 (-0.22, -0.03)	<0.01	-0.10 (-0.18, -0.02)	0.02
Unpurged − Allogeneic	0.03 (-0.06, 0.13)	0.49	0.03 (-0.06, 0.12)	0.51	0.16 (0.08, 0.24)	< 0.01
Purge − Allogeneic	0.11 (-0.02, 0.23)	0.09	0.16 (0.04, 0.27)	< 0.01	0.26 (0.15, 0.37)	< 0.01

Table 3. Estimate of survival and standard error at 500 days from initiation of therapy.

Type of Transplant	Kaplan-Meier	Direct Adjusted		Via Baseline Command	
		Stratified	Unstratified	Stratified	Unstratified
S1-S3	0.45 (0.09)	0.61 (0.06)	0.65 (0.06)	0.67 (0.08)	0.71 (0.08)
S4	0.74 (0.06)	0.67 (0.06)	0.62 (0.05)	0.74 (0.07)	0.70 (0.06)

Table 4. Estimated di erence in survival at 500 days from initiation of therapy.

Di erence	Kaplan-Meier		Stratified		Unstratified	
	Estimate (95% CI)	P	Estimate (95% CI)	P	Estimate (95% CI)	P
S1-S3 − S4	-0.30 (-0.52, -0.08)	< .01	-0.06 (-0.24, 0.12)	0.51	0.02 (-0.12, 0.17)	0.73