

The Hazards of Modeling Proportional Hazards in SAS PHREG (SAS version 8.2) vs. TPHREG (SAS version 9)

Erika Daly, ICON Clinical Research, Dublin, Ireland

Farzana Sufi, ICON Clinical Research, Marlow, UK

Aline van Maanen, ICON Clinical Research, Dublin, Ireland

ABSTRACT

In SAS® version 8.2, PROC PHREG performs regression analysis of survival data based on the Cox proportional hazards model. The syntax is similar to other regression procedures within SAS, but the predictors must all be numeric variables. There is no built-in facility for dealing with categorical predictor variables, interactions, etc. In order to fit such terms to a model, separate dummy variables must be calculated for each level of the categorical variable in question, and for each interaction to be included. We will demonstrate how this can be done in PROC PHREG in SAS v8.2. We will also show how easy it is to fit models using PHREG and believe you have the 'correct' results, even when inappropriate indicator variables have been set up.

SAS v9 contains PROC TPHREG which is currently a test procedure for PHREG. TPHREG allows the use of CLASS statements, making the fitting of such categorical predictors easier. The results from version 8.2 will be compared to version 9 to illustrate how the outputs from the various model statements differ.

INTRODUCTION

Statistical methods used for the analysis of survival data in indications such as Coronary Heart Failure, Oncology, etc must account for both censored and non-censored data. One of the best known methods is the Cox's proportional hazards model, which is a semi-parametric model. The model assumes a parametric form for the effects of the explanatory variables, but it allows an unspecified form for the underlying survivor function.

In a recent Phase III clinical trial appreciable differences were found between the parameter estimates (hazard ratios and interpretations) generated using SAS v8.2 for development and SAS v9 for validation for an interaction term. This is due to the implicit assumptions that v8.2 makes about the input variables. If these assumptions are not valid for the given data then PHREG may give inappropriate and unreliable results.

DATA

Due to client confidentiality, the actual data from this Phase III trial cannot be used. Instead, bone marrow transplant data from an online source ^[1] is used to replicate the issues found. This data has been modified to more closely resemble the Phase III data. In the original dataset, there were three disease groups, four hospitals and several indicator variables. Only two disease groups and three hospitals are included in this analysis, yielding a total of 77 patients. The variables selected are Time to Death (and its associated censoring flag), Disease Group and Hospital.

Disease group is coded as:

1 = Acute Lymphoblastic Leukemia (ALL)

2 = Acute Myelogenous Leukemia (AML) – Low risk

(Disease group 3 is omitted)

Hospital is coded as:

1 = The Ohio State University Hospital

2 = Alferd Hospital

3 = St. Vincent's Hospital

(Hospital 4 is omitted)

The dataset for this analysis is named BMT.

MODEL 1 – MAIN EFFECTS AND INTERACTIONS

CODING

The model being fitted is quite simple: can disease group (GROUP) and hospital (HOSPITAL) predict survival time in bone marrow transplant patients? In addition to these main effects, a term for the interaction between the main effects (group-hospital interaction: GHINT) is also fitted. The code to fit this model in SAS v8.2 is:

```
proc phreg data=BMT;  
    model time*status(1) = group hospital ghint;  
    ghint=group*hospital;  
run;
```

Although SAS will fit the model and provide estimates, the interpretation of these estimates is not straightforward.

OUTPUT

Analysis of Maximum Likelihood Estimates

Variable	DF	Parameter	Standard	Chi-Square	Pr > ChiSq	Hazard Ratio
		Estimate	Error			
group	1	-1.01445	0.83217	1.4860	0.2228	0.363
hospital	1	0.92289	0.77017	1.4359	0.2308	2.517
ghint	1	-0.21761	0.45123	0.2326	0.6296	0.804

There is only 1 degree of freedom for the HOSPITAL variable, and as the variable has three categories it should have 2 degrees of freedom. It is, of course, possible to derive the hazard ratios between the groups and hospitals based on these figures by taking into consideration the coding applied to the variables as outlined in the Data section above. Even still, there is no measure of the overall effect of either group or hospital on survival times. Familiarity with other regression procedures within SAS, such as PROC MIXED, give rise to the expectation that the p-values produced by this model are the p-values for the overall effect – especially as only one estimate is produced per variable – and this is where the problems with interpreting the output begin.

The discrepancy between the interpretation of the output from PROC PHREG and the output from other regression procedures is rooted in the fact that the PHREG procedure is not treating these variables as categorical variables. Thus, it is assumed that the hazard between hospital 1 and hospital 3 is twice the hazard between hospitals 1 and 2 (ignoring the interaction term for the moment). As the coding is arbitrary, such interpretations of the results are meaningless.

CREATING INDICATOR VARIABLES

CODING

SAS denotes categorical variables through the use of a CLASS statement, which has become so much of a standard part of coding with clinical trials that sometimes its importance can be overlooked. Version 8.2 has no facility for including a CLASS statement in the PHREG procedure, so indicator variables need to be created. This can be done using PROC GLMMOD.

```
proc glmmmod data=BMT outparm=BMTparm outdesign=BMTdesign;
  class group hospital;
  model time*status(1) = group hospital group*hospital;
run;
```

OUTPUT

Parameter Definitions

Column Number	Name of Associated Effect	CLASS Variable Values	
		group	hospital
1	Intercept		
2	status		
3	group	1	
4	group	2	
5	hospital		1
6	hospital		2
7	hospital		3
8	group*hospital	1	1
9	group*hospital	1	2
10	group*hospital	1	3
11	group*hospital	2	1
12	group*hospital	2	2
13	group*hospital	2	3

A dataset is created based on the BMT data with 13 additional columns (named COL1 to COL13) containing the indicator variables as described in the table above. The name of this new dataset is specified in the outdesign= option in PROC GLMMOD; in this example it is called BMTdesign. Note that as all indicator variables are included in this dataset, selecting all the columns will cause co-linearity in the model statement.

Taking “1” as the baseline level for both group and hospital, the indicator variable for GROUP is given by COL4. COL6 and COL7 are the indicator variables for hospitals 2 and 3 respectively. The indicator variable for the interaction between group and hospital 2 is given by COL12 and COL13 is the indicator variable for the interaction between group and hospital 3.

MODEL 2 – USING INDICATOR VARIABLES CODING

Using the BMTdesign dataset, fit a model containing the indicator variables for the main effects and interaction selected above.

```
proc phreg data=BMTdesign;
  model time*status(1) = col4 col6 col7 col12 col13;
  group:    test col4;
  hospital: test col6 col7;
  ghint:    test col12 col13;
run;
```

The TEST statement produces overall p-values for the groups, hospitals and interaction, labelled by the word(s) appearing before the colon.

OUTPUT

Two outputs of interest are generated from this model: the standard parameter estimates and also the overall p-values for the effects (as determined by the grouping of the indicator variables) listed in the TEST statements.

Analysis of Maximum Likelihood Estimates

Variable	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio
Col4	1	-1.26608	0.51800	5.9741	0.0145	0.282
Col6	1	1.74803	0.92867	3.5430	0.0598	5.743
Col7	1	1.54350	0.80377	3.6877	0.0548	4.681
Col12	1	0.89363	1.35993	0.4318	0.5111	2.444
Col13	1	-0.55635	0.97444	0.3260	0.5680	0.573

Linear Hypotheses Testing Results

Label	Wald		
	Chi-Square	DF	Pr > ChiSq
group	5.9741	1	0.0145
hospital	5.0437	2	0.0803
ghint	0.9581	2	0.6194

COMPARISON OF MODEL 1 AND MODEL 2

Parameter Estimates, and hence the interpretation of the results, are very different between Model 1 and Model 2. For example, in Model 2 there is a significant difference in hazard of dying between patients in Group 1 and Group 2 ($p=0.0145$). In Model 1, the (incorrect) hazard ratio seems to suggest that the difference between groups is non-significant ($p=0.2228$). Of course, these ratios have to be adjusted for the presence of the interaction term, but nonetheless it illustrates the possible pitfalls in interpreting the output if the model has been incorrectly specified.

MODEL 3 – VERSION 9 AND TPHREG

A new, enhanced version of PHREG is currently available in SAS version 9. TPHREG, which is undergoing validation (“T” stands for “test”), allows the addition of a CLASS statement before the MODEL statement. This enables SAS to automatically create the necessary indicator variables for Cox’s regression, similar to other regression procedures within SAS such as PROC GLM and PROC MIXED.

```
proc tphreg data=BMT;
  class group hospital;
  model time*status(1) = group hospital group*hospital;
run;
```

The output from PROC TPHREG yields identical parameter estimates to those obtained from Model 2 (version 8.2), where the categorical variables were manually defined.

CONCLUSION

In the original Phase III trial that highlighted this problem, the primary analysis model included main effects and an interaction term, with the interaction term to be dropped if it was non-significant. When the incorrect Model 1 was fitted, the p-value for the interaction term was 0.260, and so it was concluded that the term should be dropped and the primary analysis based on a model with main effects only. However, fitting the indicator variables model (Model 2) gave a different story: the p-value for the interaction term was 0.060 (significant at the 10% level) and needed to be included in the primary analysis model. Naturally, the omission of the interaction term from a model when it was required greatly impacted on the interpretation of the model. The conclusions drawn regarding the efficacy of the trial drug being examined were also affected.

TPHREG offers an attractive alternative to manually creating the required indicator variables. However, as TPHREG is currently in testing, its output should still be regarded with some caution. In some cases, it was noted that TPHREG produced missing values for the Wald chi-square and p-values, though using indicator variables in v8.2 provided the parameters required.

Though the same information regarding the effect of switching between hospitals or groups on survival time can be obtained from Model 1 as from Model 2 or 3, the construction and use of indicator variables allows for a more intuitive interpretation of the output presented. Only once the correct model has been fitted should any decisions regarding the propriety of including the interaction term be made. Only then can the true interpretation of the data be made.

REFERENCES

^[1] Source data can be found at <http://www.biostat.mcw.edu/homepgs/klein/bmt.html>

CONTACT INFORMATION

If you have a question or would like to discuss this paper, please contact the author at:

Erika Daly,
ICON Clinical Research,
South County Business Park,
Leopardstown,
Dublin 18,
Ireland.

Or by email at:

DalyE@iconirl.com