Survival Analysis of the Data of Recurrence for Kidney Patients

SUMMARY OF FINDINGS

(1) A Cox model with terms for *gender* and *pkd* is established via backward model selection for the data of effect using the portable dialysis machine on kidney patient recurrence. The adequacy of model has been assessed and confirmed.

(2) The gender and pkd type of disease shows significant effect on the recurrence of patients with kidney disease in the current model. Controlling other covariates, male patients are 1.607 times more likely to experience infection than females; and patients with disease of pkd type have 8.053 times more risk to experience infection than that of patients without the disease of pkd type.

(3) In the presence of the other covariates, the age does not appear a strong association with infection, and the disease of type gn and the disease of type an are insignificant at predicting the survivorship with the significance of α at 0.05. The infection risk for older patients to use the dialysis machine is estimated to be higher than that of a younger patient. For those patients with either the type gn disease or type an diseases have less chances to get infection to use the dialysis machine, than that of those patients with none of the two type diseases.

STATISTICAL ANALYSIS

Preliminary Diagnostics (**Please refer to the appendix for some of the statistical analysis' figures.)*

Univariate analysis is carried out to identify the effect of individual variable on time to the recurrence of kidney patients. For the categorical variables (gender and disease of type an, type gn and type pkd) , the non-parametric model KM curves are generated (as shown in Figure 1), and the significance of impact is tested by log-rank test individually, and summarized in Table 1. Cox proportional hazard regression is performed to check the continuous variable (age10, **age10=age-10, age10 instead of age will be used for the following analysis.*).

Basing on the plots of KM, the curves for the levels of "gender", "an" and "pkd" are different from each other, which indicates effect of sex and effect of disease types (pkd and an) might exist (Figure 1). There is some cross-over in the curves for levels of "an", which implies that assumption of proportional hazard rate could be challenged. For the curves of "gn", there is not much difference for the two levels, indicating that gn might not have much impact on the infection risk for patients using the dialysis machine, which is also supported by the result of log-rank test with a large p-value at 0.6791 (Table 1).

Table 1

Strata Variable	gender	gn	an	pkd
Pr>Chi_Sqr	0.3957	0.6791	0.0860	0.1001

The result for Cox proportional hazard regression shows that p-value of "age10" is "0.9958", indicating that age might not be strongly associated with the infection risk for patients using the dialysis machine.

Full First-order Model:

Methods of Breslow, Efron and exact are used to fit the Cox proportional hazards model containing all covariates using for handling tied survival times. The Efron method is selected to be used for the following analysis since Efron's method approximation is adequate to use and is less computationally demanding (Table 2).

Table 2

Parameter	Breslow		Efron		Exact	
	Estimate	**S.E.**	**Estimate**	**S.E.**	**Estimate**	**S.E.**
age10	-0.00155	0.01612	-0.00169	0.01613	-0.00170	0.01613
gender(=0)	0.96067	0.49802	0.96595	0.49847	0.96635	0.49864
gn(=0)	-0.07339	0.58533	-0.07179	0.58530	-0.07150	0.58537
an(=0)	-0.81794	0.62707	-0.81242	0.62668	-0.81230	0.62668
pkd(=0)	2.07987	1.19660	2.08346	1.19633	2.08371	1.19635

According to the output of SAS using efron method (Table 2), the fitted full first-order Cox model is as following:

$h_{\underline{x}}(time1)=h_{\underline{0}}\ time1^*\ \exp(-0.00169*age10 + 0.96595*gender - 0.07179*gn - 0.81242*an + 2.08346*pkd)$,

where *time1* is the failure time in days; age10 is adjusted patient's age by subtracting 10; gender= 1 if sex is female and = 0 if sex is male; *gn*= 1 if Disease type is GN and = 0 otherwise; *an*= 1 if Disease type is AN and = 0 otherwise; *pkd*= 1 if Disease type is pkd and = 0 otherwise.

**The baseline patient is someone with age=10, gender = female, gn= yes, an= yes, pkd = yes*.

All of the terms in the initial model are insignificant, the terms of gender and pkd approaching significance (data not shown). The different interactions in the model are also checked, but no significant ones exist (data not shown).

Residual Analysis for the Full First-order Model:

Plots of Martingale residuals against age10 (Figure 2) show LOESS curves with approximated slopes of 0 and intercepts of 0, so no transformations on the continuous covariates are needed. However, there is somewhat non-linear after age10 at 22, so creating an indicator function at age10 less than 22 (the turning point). But the age is still not significant in the model even after adding this indicator to the model (data not shown).

Almost 29% of the observations in the dataset are censored. The plot of deviance residuals vs. risk age (Figure 3) does not show departure from normality with no extreme outliers beyond the ± 3 limits so the overall fit is reasonable.

Check for Proportional Hazards Assumption:

The plots of scaled Schoenfeld residuals vs. failure time for age10 does not show obvious time-dependent patterns, although the plots for age10 vs. time (Figure 4) show somewhat increasing trends not centered at 0. Besides, OLS regression of the Schoenfeld residuals versus time indicates a linear fit with roughly zero slope and zero intercept. These predictors appear to satisfy the proportional hazard assumption of Cox model. Hence, age does not seem to be a time-dependent variable here.

The proportional hazard assumption was also checked for the other indicator variables (Figures not shown) and they appear also to satisfy this condition except for the variable, an.

Backward Model Selection:

Since there is no significant help to add one adjusted variable about age in the previous analysis of proportional hazards assumption, the original dataset will still be used for model selection.

A backward model selection (Output 1) is conducted with a 5% significance level for removing effect in each step. Age10, gn and an with p-values bigger than $\alpha = 0.05$, are removed step by step until all the remaining terms of *gender and pkd*, are significant to fit the model.

Output 1

Analysis of Maximum Likelihood Estimates

Parameter		DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio	Label
gender	0	1	0.95825	0.47500	4.0698	0.0437	2.607	gender 0
pkd	0	1	2.20308	1.10327	3.9875	0.0458	9.053	pkd 0

Summary of Backward Elimination

Step	Effect Removed	DF	Number In	Wald Chi-Square	Pr > ChiSq
1	age10	1	4	0.0110	0.9165
2	gn	1	3	0.0081	0.9283
3	an	1	2	2.2665	0.1322

Output 2

Analysis of Maximum Likelihood Estimates

Parameter		DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio	Label
gender	0	1	0.95825	0.47500	4.0698	0.0437	2.607	gender 0
pkd	0	1	2.20308	1.10327	3.9875	0.0458	9.053	pkd 0

Output 3

Fitted Model

Model Fit Statistics

Criterion	Without Covariates	With Covariates
-2 LOG L	138.910	131.768
AIC	138.910	135.768
SBC	138.910	138.359

Initial Model

Model Fit Statistics

Criterion	Without Covariates	With Covariates
-2 LOG L	138.910	129.492
AIC	138.910	141.492
SBC	138.910	149.267

Output 4

Supremum Test for Proportionals Hazards Assumption

Variable	Maximum Absolute Value	Replications	Seed	Pr > MaxAbsVal
gender0	0.9868	1000	1510930839	0.1830
pkd0	0.5976	1000	1510930839	0.3640

Fitted Final Model: *(* global tests for this final model are very significant at α=5% respectively)*

Basing on the output of SAS (output 2), the fitted final Cox model:

$h\underline{x}(time1)=h\underline{0}$ time1 exp(0.95825*gender + 2.20308*pkd).

A partial F test that tests is conducted to assess the final model's adequacy of fitting: H0: $\beta_{age10} = \beta_{gn} = \beta_{an}$ vs. H1: Not all βxs are equal zero. Using the model fit statistics for the final model and for the initial model (Output 3), the test statistic is calculated to be

Λ=-2(logL0 – logL1) = 131.768-129.529=2.239, which is smaller than the critical point that is $\chi^2_{\alpha=0.05,df=3} = 7.815$, then we fail to reject H_0 at α = 0.05 and conclude that the data do not provide sufficient evidence to support that the final model fits data significantly worse than the initial model does.

The proportional hazard assumption is also checked for the fitted model. All the variables (gender and pkd) seem to satisfy the proportional assumption for Cox model (Figure 5 & Output 4).

Interpretations Based on the Final Model:

(1) *gender* (p-value = 0.0437) is significant at predicting the survivorship in the presence of the other covariates at α = 0.05. Controlling for all other covariates, the risk of recurrence for a male patient using the dialysis machine is estimated to be 1.607 times of risk to experience recurrence higher than that for a female patient under the same condition with a 95% Wald confidence interval that contains 1.

(2) In the presence of the other covariates at α = 0.05, the *pkd* type of patient (p-value=0.0458) is 9.053 times of that for a female patient without pkd disease with a 95% Wald confidence interval that contains 2.

(3) Age and the other two types of diseases (gn and an) are insignificant and removed from the current model.

DISCUSSION

This study is conducted by fitting the survival data using a Cox model. Based on this model, gender and the pkd type of disease show a significant impact on the recurrence of kidney patients using the current dialysis method. However, as Figure 1 shows, the KM estimates of survival curves for the genders cross over after 300 days, indicating that the effect of sex might not be constant through the failure or censoring times. A model that stratifies the failure or censoring times can be utilized to investigate this issue further. It was shown by the published papers and research analysis reports, age has an important correlation with the recurrence of kidney patients. However, in current case, the age does not show any significant association with risk of infection, after adjusting for gender and type of disease. The reasons for this might result from small sample sizes or more complicated associations or interactions that are not involved. For our model, we did not separate random effect for each patient. If considering this factor, we could try gamma frailty model to fit to the data.

APPENDIX

Figure 1

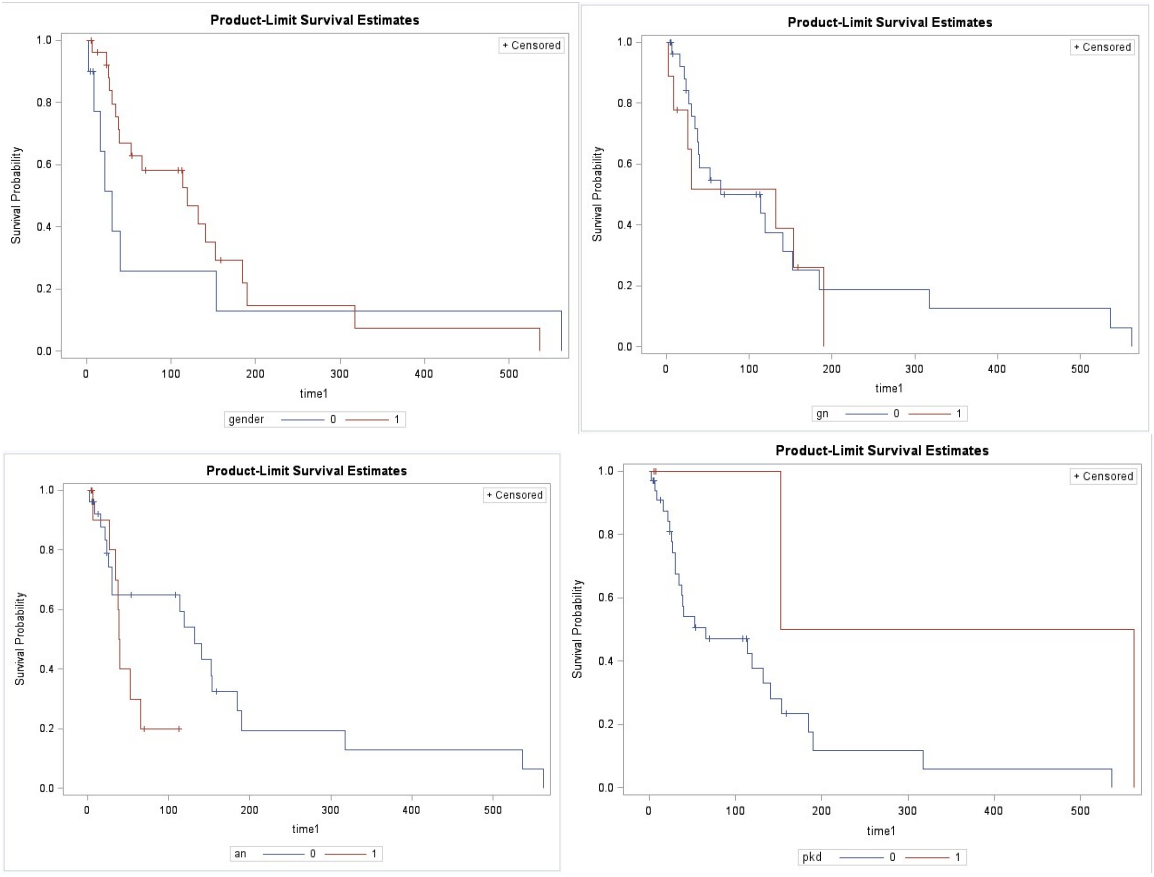

Figure 2

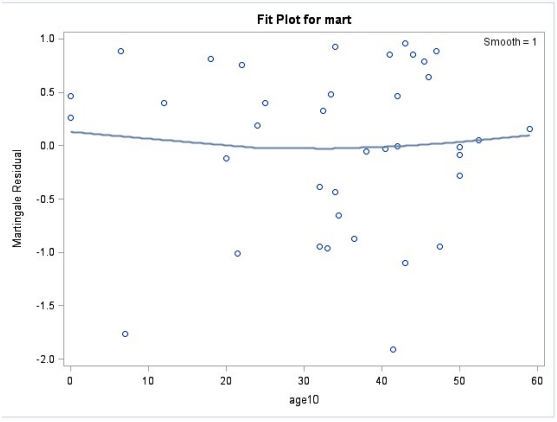

Figure 3

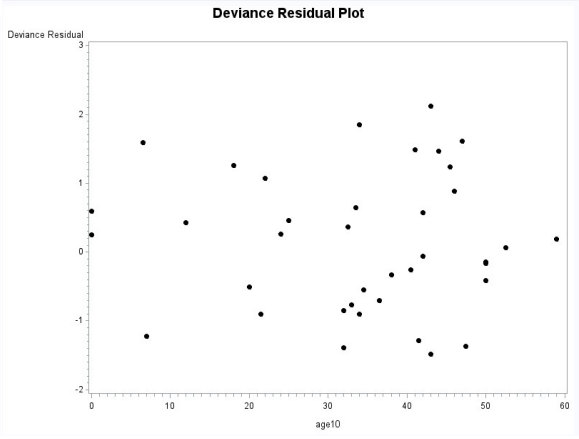

Figure 4

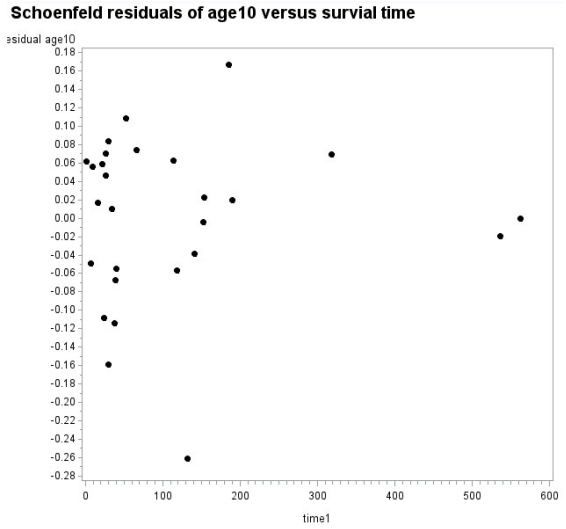

Figure 5