Department of Biostatistics
Erasmus University Medical Center

Internal data analysis Report

Title	PRIAS
Version	1
Date	31 January 2017
Author	
Collaborators	

Visiting address: Wytemaweg 80, 3015 CN Rotterdam, Na building (room Na-2418)
Postal address: PO Box 2040, 3000 CA Rotterdam, the Netherlands
Web: http://www.erasmusmc.nl/biostatistiek/

1 Introduction

In this report we present the analysis of the prostate cancer data set from the PRIAS study. The original data set as of December 7 2016 contains contains information about 6039 patients who received kidney transplants. PSA, DRE and Gleason scores were obtained from these patients based on a fixed schedule. The age of the patients at baseline was also available. The goal of the analysis we present in this report was to know how do the 3 aforementioned biomarkers affect time to progression of cancer and how are they related to each other. To perform the analysis, we only considered 5163 patients out of the 6039 available. The rest of the patients were not considered because of the following reasons.

- Patients having no other information other than the ID.
- Patients having age less than 5.
- Patients not having any information on follow ups.
- Patients who discontinued but do not have any date of discontinuation.
- Patients not having reasons of discontinuation.
- Patients with incomprehensible reasons of discontinuation.
- Patient with ID 947, because he did not have any Gleason measurements and thus a multivariate analysis with this patient was not possible.

Further we also removed all those observations where:

- PSA/DRE/Gleason scores were available but the date of measurement was not.
- Date of measurement is available but neither of the 3 scores are available.
- Measurements were dummy.
- Gleason scores were 0 or 1.

For the patients who did not have an event till the last follow up date, we considered them to be lost to follow up (i.e. censored) and their last follow up date was considered as the date of censoring. Further we combined the categories, disease progression or death from cancer, into the event category and all other categories (anxiety, watchful waiting, death from other reasons, lost to follow up) into the censored category. Lastly, we dichotomized DRE and Gleason scores because some of the categories of Gleason and DRE scores had too less patients and besides our software doesn't support categorical outcomes with more than 2 levels. We considered Gleason scores which were less than or equal to 6 as low and the greater than 6 as high. For DRE, all DRE scores higher than T1c were considered high and the rest as low.

2 Time to event analysis of disease progression

We used a Cox model for time to event analysis of disease progression. Out of the 5163 subjects considered for analysis, 1178 had the event of interest and the rest were censored. The only covariate available for analysis was Age, and we found both the linear and quadratic effect of Age to be significant.

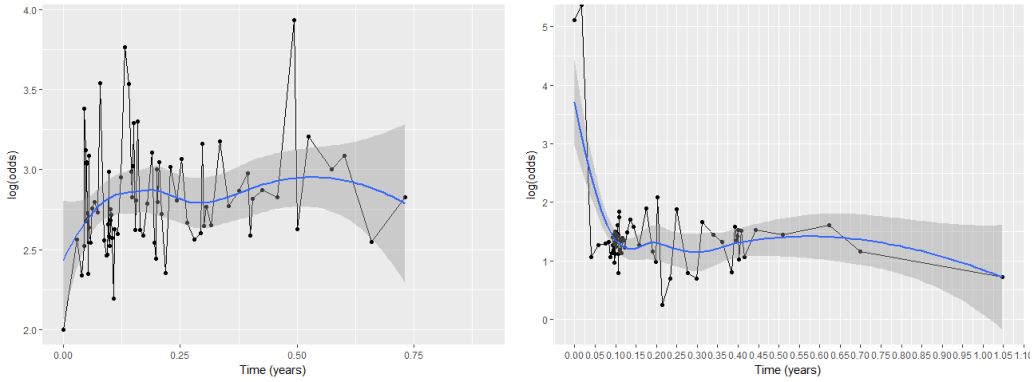
3 Multivariate longitudinal analysis of PSA/DRE and Gleason

Although we could not complete the multivariate longitudinal analysis of the 3 outcomes, we present the steps we took for univariate longitudinal analysis.

For the longitudinal analysis of DRE and Gleason scores, we first plotted the observed evolutions of the log odds based on a categorized time scale. This was done because we will model the log odds for both of these responses using Generalized linear mixed models (GLMM). The

observed evolutions for log(odds) are shown in Figure 1a and Figure 1b. Based on these plots we decided to use spline with 1 non boundary knot to model the evolution for both of these responses. For the random effects we only considered the random intercept for both of the responses.

In the case of PSA we observed that certain patients had very high levels of PSA and some had PSA 0. This indicated that PSA scores can have a right skewed distribution. We thus transformed these scores using $\log(PSA + 1)$ transformation. We then checked the trend plots and did not observe any clear marginal trend for the patients (Figure 2). However on checking the individual trends, we observed that they varied quite a lot from subject to subject. After comparing various models based on fitted PSA scores for patients, and considering the computational restrictions we chose a longitudinal model having a spline with 3 internal knots to model fixed effect and with 1 internal knot to model random effects.



(a) Observed evolution of log(odds) for DRE score.

(b) Observed evolution of log(odds) for Gleason score.

Figure 1: Observed evolutions of log(odds) for DRE and Gleason scores. The time scale is categorized. Thus log(odds) at time point t are calculated by taking all observations between time t and $t-1$.

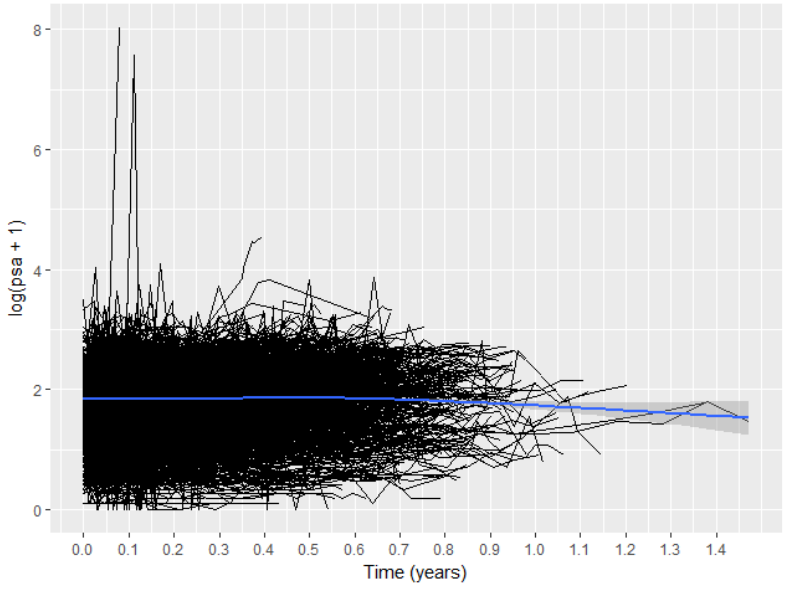


Figure 2: Observed evolutions for PSA measurements. The PSA measurements are transformed as log(psa + 1).

Variable	Mean	Std. Dev	2.5%	97.5%	P
Intercept	-7.430	0.788	-9.043	-6.102	0.000
Age	0.048	0.008	0.035	0.065	0.000
Spline (visitTimeYears): 1	4.830	0.710	3.679	6.224	0.000
Spline (visitTimeYears): 2	-7.311	0.927	-9.185	-5.640	0.000

Table 1: Longitudinal submodel estimates for joint model with Gleason score outcome. Reference outcome is Low gleason score.

Variable	Mean	Std. Dev	2.5%	97.5%	P
Age (orthogonal polynomial)	-29.397	7.378	-43.827	-15.373	0.000
Age * Age (orthogonal polynomial)	-6.604	4.611	-15.634	2.686	0.148
odds (High against Low)	1.947	0.206	1.573	2.361	0.000
Slope (odds)	0.491	0.128	0.223	0.723	0.002

Table 2: Survival submodel estimates for joint model with Gleason score outcome.

4 Joint model for PSA, DRE and Gleason scores

Since we could not fit a multivariate longitudinal model, we present 3 separate joint models, consisting of longitudinal submodels for PSA, DRE and Gleason measurements. The models were fitted with mvJointModelBayes API instead of jointModelBayes because the jointModelBayes API did not lead to models with convergence for MCMC outcomes. Convergence was evaluated graphically using traceplots.

4.1 Gleason score

In the joint model for Gleason score, the MCMC outcomes did not show convergence for the random intercept variance component. However for the survival submodel parameter estimates did not show signs of non convergence, including the ones for baseline hazard. The resulting parameter estimates are shown in Table 1 and Table 2. It can be seen that both odds and the slope odds are strongly associated with time to event. More specifically, if at any given time point t, a person has twice the odds of getting a High gleason score in biopsy then what he has now at t then his hazard of having progression (time for intervention via treatment/death due to cancer) at the same time point become 7 times of original hazard. More imporatantly, for this scenario we have to assume that the rate of change of odds of biomarker is the same in either scenario and the patient is also the same.

4.2 DRE score

In the joint model for DRE score, the MCMC outcomes did not show convergence for the random intercept variance component. However the non convergence was not as serious as in the case of Gleason score. The rest of the parameters did not show signs of non convergence. The resulting parameter estimates are shown in Table 3 and Table 4. It can be seen that only the of slope of odds is associated with time to event. This means that if, for a patient at any given time point t the rate at which odds (of Low against High) change become twice than what he has now, then his hazard of having progression (time for intervention via treatment/death due to cancer) at the same time point are 12 times of original hazard. This however seems quite counterintuitive because high DRE scores indicate a more severe form of cancer. Nonetheless, once again we have to assume that the patient remains the same in either of the two scenarios for the interpretation to stay valid.

4.3 PSA score

In the joint model for PSA score, the MCMC outcomes did not show any sign of non convergence for any of the parameter estimates including the components of variance covariance matrix. The resulting parameter estimates are shown in Table 5 and Table 6. It can be seen that both log(PSA + 1) and its slope are associated with time to event. To interpret this let us assume a patient

Variable	Mean	Std. Dev	2.5%	97.5%	P
Intercept	-9.732	0.655	-11.023	-8.504	0.000
Age	0.078	0.009	0.061	0.095	0.000
Spline (visitTimeYears): 1	-1.641	0.400	-2.397	-0.862	0.000
Spline (visitTimeYears): 2	3.144	0.938	1.292	4.867	0.000

Table 3: Longitudinal submodel estimates for joint model with DRE score outcome. Reference outcome is T1c DRE score.

Variable	Mean	Std. Dev	2.5%	97.5%	P
Age (orthogonal polynomial)	3.507	4.878	-3.566	15.021	0.486
Age * Age (orthogonal polynomial)	-3.878	3.982	-12.932	1.297	0.304
odds (T1c against >T1c)	0.099	0.078	-0.034	0.253	0.214
Slope (odds)	-2.548	0.677	-3.596	-0.972	0.002

Table 4: Survival submodel estimates for joint model with DRE score outcome.

at some time point t where his PSA score is 4. If instead this PSA score would've been almost twice (actual is $2 \times \text{PSA} + 1$) then his hazard of having progression (time for intervention via treatment/death due to cancer) at the same time point would've been 1.3 times of original hazard. We assume that the patient here can be different while doing the comparison, however the rate of change of log(PSA + 1) remains the same in either scenario.

Variable	Mean	Std. Dev	2.5%	97.5%	P
Intercept	1.850	0.005	1.841	1.860	0.000
Age (orthogonal polynomial)	1.590	0.358	0.889	2.259	0.000
Age * Age (orthogonal polynomial)	-1.707	0.355	-2.389	-1.026	0.000
Spline (visitTimeYears): 1	0.075	0.007	0.062	0.088	0.000
Spline (visitTimeYears): 2	0.278	0.020	0.240	0.315	0.000
Spline (visitTimeYears): 3	0.132	0.073	-0.007	0.276	0.058
Spline (visitTimeYears): 4	0.069	0.140	-0.201	0.334	0.606
Residual S.D.	0.189	0.001	0.188	0.191	0.000

Table 5: Longitudinal submodel estimates for joint model with log(PSA + 1) score outcome.

Variable	Mean	Std. Dev	2.5%	97.5%	P
Age (orthogonal polynomial)	7.203	2.754	1.868	12.590	0.012
Age * Age (orthogonal polynomial)	-4.101	2.836	-9.849	1.590	0.132
log(PSA + 1)	0.414	0.106	0.204	0.627	0.000
Slope log(PSA + 1)	0.688	0.063	0.564	0.814	0.000

Table 6: Survival submodel estimates for joint model with log(PSA + 1) score outcome.

References