

Nephropathy prognosis in DK and HK

BxC, September 2014

Compiled Tuesday 23rd September, 2014, 18:02
 from: C:/Bendix/Steno/HongKong/protocol1.tex

Contents

1	Introduction	1
1.1	Definitions and abbreviations	2
2	Trajectories in clinical variables prior to ESRD	2
2.1	Latent class trajectories	2
2.1.1	Reporting results	3
2.1.2	Choice of response variable	3
2.1.3	Current status of analyses	3
2.2	Joint models	3
2.3	Data requirements	4
3	Clinical course of patients after DKD	4
3.1	Prediction from baseline	4
3.1.1	Data requirements	4
3.2	Updated clinical measurements	4
3.2.1	Data requirements	4
3.3	Joint modelling	5
4	Clinical measurements	5
5	Documentation	5
6	Collaboration	5
	References	6

1 Introduction

This is an updated protocol exclusively dealing with the substance of the proposed analyses and with very limited attention to the details of the clinical variables required, let alone available. It is updated with the results of conversations between Bendix Carstensen and Ronald Ma at the EASD in Vienna 2014.

1.1 Definitions and abbreviations

CKD Chronic kidney disease, also in the literature referred to as DN - diabetic nephropathy or NRA - nephrotic range albuminuria. In this study defined as $\text{eGFR} < 60 \text{ mmol/min}$

DKD CKD in a diabetes patient.

DN Diabetic nephropathy defined from (micro/macro) albuminuria ($> 30, 300\text{mg}/24\text{h}$) or albumin/creatinine ratio in spot urine ($> 250 \text{ mg/g}$ (M), $> 350 \text{ mg/g}$ (F)).

ESRD End stage renal disease - kidney failure, defined as either dialysis, transplant or $\text{eGFR} < 15 \text{ mmol/min}$ or death with mention of kidney disease.

2 Trajectories in clinical variables prior to ESRD

2.1 Latent class trajectories

This is trajectory analysis looking back in time from diagnosis of ESRD. That is, the time-development of the variable of interest (eGFR or uAlb) is modelled using for example cubic splines, attempting a grouping of persons into classes with a common shape and location of the curves. For persons who progress to ESRD, the curves can have any shape as a function of time before ESRD because the time-scale is well-defined from the origin (date of ESRD). However, for persons who do not have an event, the time-scale does not have an origin, and so any model for the response should be invariant under *individual* shift of origin of the timescale.

Hence the practical approach is to do the latent class trajectory models with a specified class of models for the shape of the response variables *only* for the patients that progress to ESRD. Subsequently, persons are assigned to their most likely latent class and the model refitted for the data with *fixed* (latent) classes including also the non-ESRD patients with a specified linear effect of the time-scale and necessarily a random intercept and random slope for these.

The latter is because the individual assignment of time-scale for the non-ESRD patients is arbitrary, and reshuffling it between them will change the distribution and in particular the between-person variance. If there is not an individual linear (random) slope with time for each individual, the *mean* in the non-ESRD group is not guaranteed to be linear, because of the different allocation of persons over the timescale. Hence, a post-hoc model as the one outlined above should allow for both random effects and residual variances to be different between the ESRD and non-ESRD persons.

Notably, the residual variance and the between-person random effects variances from the non-ESRD group should be considered nuisance parameters, since they are dependent on the allocation of the “time before EASD” in the non-ESRD group. It should be possible to get a handle on the likely range of the “true” variance components in the non-ESRD group by repeating the analysis on datasets where the timing of the measurements in the non-ESRD group have been perturbed. The question is however how to choose the size of the perturbations.

There is actually neither any particular reason to assume that the between-person variation is the same in all latent classes, so the modelling should allow for separate variances between latent classes. This actually facilitates (some) analyses, as they will simply be separate analyses of each latent class.

2.1.1 Reporting results

The result of this type of analyses is a grouping of patients in latent classes (that is for each person a posterior probability of belonging to each of the latent classes), as well as an estimated (mean) eGFR-curve for each of the classes.

To the extent that the class discrimination is good — that is the two largest probabilities for a patient are not too close — it is of interest to use the patients' characteristics *within* each latent class to describe the characteristics of the class. Basically, using the clinical variables to predict (*i.e.* describe) the latent classes.

Also, it would be desirable to do the post-hoc analyses on the subsets of patients with posterior probabilities for the most likely latent class above different thresholds, say 50, 60, ..., 90% to see if substantially different mean patterns emerged. One would expect to see decreasing within-class variances by increasing threshold.

2.1.2 Choice of response variable

If the endpoint is ESRD, the most interesting analysis-variable (response) is uAlb, and only to a lesser extent eGFR, since the latter is a part of the event-definition (eGFR < 15 mmol/l is a sufficient condition for ESRD).

The required data for this type of analysis is therefore these two variables together with dates of the measurements. And of course the relevant clinical measurements at these dates too.

2.1.3 Current status of analyses

This type of analysis has already been undertaken partially in Hong Kong for (log?) eGFR-measurements based on local PWH-HK data, except that the preliminary analyses were using time since entry. They should be re-done using time before ESRD, discarding any information post-ESRD, and also expanding the analysis to uAlb as response variable.

Similar analyses will be done on the SDC data, identifying mean trajectories in 2, 3 and 4 classes among ESRD patients.

2.2 Joint models

Joint models for clinical parameters and occurrence of DKD (looking forward in time), are models where a random effect (or latent class) *both* influences the trajectory of the clinical measurements (eGFR, primarily, maybe also BP), *and* the occurrence of ESRD.

2.3 Data requirements

The data requirement for both these approaches would be relevant demographics and the date and value of the clinical measurements (uAlb, eGFR, other clinical measurements — see below) as well as the date of the DKD, ESRD, death and end of follow-up. The latter is the date of the last clinical measurement.

3 Clinical course of patients after DKD

Another area of particular interest is the description of the clinical course of patients through states of complications, and major other events, such as CVD and ESRD. This can also be approached in different ways:

3.1 Prediction from baseline

A pure prediction model for complications course as a multistate model with transitions through different levels of complications, but using only clinical baseline values to predict the course. This will enable prediction based on baseline values only; so this is a classical type of prediction model.

This approach allows predictions of the clinical course of patients with a given set of values of the baseline values as in [2, 1].

3.1.1 Data requirements

Baseline clinical measurements and dates of birth, DM, DKD, CVD (more than one event or type?), ESRD, death and end of follow-up (last day where status is known).

The data for this analysis would be assembled at Steno,

3.2 Updated clinical measurements

A model for the same transitions through complication states, but using *all* clinical measurements. Sometimes this is called a time-updated analysis.

This approach does however not allow prediction of individual clinical course of disease because there is no model for the time-development of the clinical variables. The major difference to the first approach is in the interpretation of the effects of the clinical parameters on rates, the first instance is the long-term effect of the baseline covariates, whereas this approach is short-term effect of covariates.

3.2.1 Data requirements

Date of each clinical visit, the clinical parameters at this date, and dates of birth, DM, DKD, CVD (more than one event or type?), ESRD, death and end of follow-up (last day where status is known)

3.3 Joint modelling

A joint model for the clinical measurements and the complication states of the patients, where *both* the (joint) trajectories of the clinical variables and the transitions between complications states are modelled.

The third, most general, approach allows in principle dynamic predictions of *both* the clinical variables *and* the complication states [4, 3]. This is the only type of model that will allow predictions forward in time given the status at *any* time after baseline, which in practice is what is desirable.

4 Clinical measurements

In the above “clinical measurements” have been used without further qualification. The clinical measurements that will be used as continuous predictors are:

HbA1c

Blood pressures

Lipids

Serum creatinine

Urine albumin

uAlb/uCr ratio

...

5 Documentation

As the project is multinational (multi-center) by nature it is important that all analytical approaches be properly documented.

Hence all analyses should be described and *both* code and results should be available *in extenso*. This means that some sort of documentation system like R-markdown (`.rmd`-files), R-weave or knitr (`.rnw`-files) should be maintained to document the analyses.

6 Collaboration

Steno shall look into the possibility to host Guozhi Jiang for some weeks in the spring of 2015. Likewise we will seek to arrange a visit by Danish collaborators to PWH in Hong Kong.

References

- [1] G. Andresdottir, M. L. Jensen, B. Carstensen, H. H. Parving, P. Hovind, T. W. Hansen, and P. Rossing. Improved prognosis of diabetic nephropathy in type 1 diabetes. *Kidney Int.*, Jun 2014.
- [2] G. Andresdottir, M. L. Jensen, B. Carstensen, H. H. Parving, K. Rossing, T. W. Hansen, and P. Rossing. Improved Survival and Renal Prognosis of Patients With Type 2 Diabetes and Nephropathy With Improved Control of Risk Factors. *Diabetes Care*, Mar 2014.
- [3] Rizopoulos D. Dynamic predictions and prospective accuracy in joint models for longitudinal and time-to-event data. *Biometrics*, 67:819–829, 2011.
- [4] Rizopoulos D. and Ghosh P. A bayesian semiparametric multivariate joint model for multiple longitudinal outcomes a time-to-event. *Statistics in Medicine*, 30:1366–1380, 2011.