Reviewer's report

Title: Optimal and continuous anaemia control in a cohort of dialysis patients in Switzerland

Version: 1 Date: 29 April 2008

Reviewer: Dorothea Nitsch

Reviewer's report:

Statistical/epidemiological review

‘Optimal and continuous anaemia control in a cohort of dialysis patients in Switzerland’

Mathieu CM et al.

I first explain what from an epidemiological perspective would be done to address the study aims. This is to clarify why this paper in its current form is problematic. I then explain how the paper could be revised. Although this means more work I do believe that the authors can address all points raised and that this ultimately may improve the paper considerably.

This paper had as main aim (outlined in the abstract):

‘To examine the control of anaemia over continuous long-term period in Switzerland’

Further at the end of the introduction there are following aims/questions:

(i) To study if the control of anaemia has improved since previous assessments in dialysed patients in Switzerland

(ii) Whether the control of anaemia and its treatment parameters could be maintained over a long period of time

(iii) Which factors modulate Hb level and Epo dose?

The first step in tackling these questions would be to define a cohort of patients, and to think carefully about internal and external validity of the study findings.

From patient perspective the ideal study design would be to assemble an incident cohort of patients. With ‘incident’ I mean following a patient from start of their dialysis treatment onwards. The findings from such an incident dialysis cohort enable the doctor to tell an individual who starts dialysis what happens for an individual with a given primary renal disease and comorbidities in terms of anemia control in the future.

From the perspective of management/hospital administrators average values in a prevalent cohort are of main interest. In different words: How do we achieve a good average Hb in the current patient population with a minimal effort? With ‘prevalent’ I mean a snapshot of all patients on dialysis, irrespective of when they
started dialysis. The problem with a prevalent cohort however is that within the first 3 months of dialysis the survival patterns of patients on dialysis tend to be completely different from what happens after 3 months. For example, in the UK diabetic people on dialysis when compared to other patients on dialysis tend to survive much better in the first 3 months than in the subsequent time, but afterwards they die much faster – particularly if they are young. Hence, survival analyses of a prevalent cohort in the UK will be affected by survival bias because a diabetic person who starts dialysis is different from a diabetic person who survived 2 years of dialysis whilst the statistical model assumes that they are identical because they are both coded as having diabetes. It is unknown whether such phenomena occur also in Switzerland. Anyway, in this study the authors were mainly interested in hemoglobin and not survival per se (see aims above). Data from the US won’t help, as these are only collected from 3 months onwards, anyway.

So, from a patient perspective one would want an incident cohort, from a managerial perspective a prevalent cohort. Afterwards one would think of an adequate sampling scheme to get a representative sample of either incident or prevalent patients on dialysis in Switzerland. Ideally one would use random sampling, as then there is no selection bias. To verify whether there is a selection bias one could then compare the data of included patients with the data on all Swiss dialysis patients held by the Swiss Renal Registry in Bern which by now should have reached good coverage.

Afterwards one would follow the patients up (according to the graph that the authors provide in figure 1), and if somebody drops out on the way one would not exclude them because of incomplete follow-up but use the data that are available. There is a vast amount of statistical literature on the bias introduced by carrying the last observation forward (LOCF) – this is data generation where there are no data. Typically LOCF data lead to more positive associations and more stability over time than there is in truth. Using LOCF is therefore not an option to deal with missing data.

After finishing the data collection a first step of analysis would be to compare those people with incomplete follow up to those with complete follow up. This is to check whether the data tell reasons for loss of follow up. Also, one would run a survival analysis to examine the association of survival with last hemoglobin measurement in those who died with those who didn’t. Reverse causality in the present data would mean that the hemoglobin measured in the months preceding death is lower because the patient was dying, and not because Swiss doctors were not good at prescribing Epo/Ferritine. For analyses with aims similar to those stated by the authors, the UK renal registry tends to run two sets of analyses - one using the full data and one where the last measurement for patients who died subsequently is deleted to see how much reverse causality may affect the conclusions.

If the aim of the study was to assess which factors affect hemoglobin control in a given patient over time in the incident cohort, then the analysis of the repeated measurements of hemoglobin would need to take account of the correlations
between measurements in a given patient, and also of the clustering of patients in dialysis centers. Analyses would model the trajectories of hemoglobin per patient as a function of patient and center factors. In an incident dialysis patient cohort time has typically a non-linear effect on hemoglobin – first hemoglobin increases rapidly, and then it remains relatively constant. There is an added problem in that erythropoietin dosage is related both to the previous erythropoietin dose and to the current levels of hemoglobin. This cannot be modeled with conventional statistical approaches. For the case of Epo dose and hemoglobin over time true multivariate models need to be used. With multivariable models I mean models that have many X values (many exposure or explanatory variables) explaining the outcome variable, with multivariate models I mean models that model many Y (outcome) variables as correlated outcomes. The effect of other variables on outcome(s) would then be assessed in univariable models (one explanatory variable), and finally in multivariable models, with a clear description of the modeling strategy leading to the final model (i.e. what variables went into the model) including some indication on how well the statistical model fitted the data. Missing values would be automatically dealt with by using expectation maximization routines in multilevel/structural models. Displays of p-values are not meaningful unless the study was actively powered to detect an effect of a given size – rather 95% confidence intervals. I guess the other reviewer who requested a statistical review of these data was after such an approach, but given that these data derive from a prevalent cohort I wonder whether any of these very complex statistical models are particularly meaningful in the setting of this particular study.

In an analysis of average values of hemoglobin at a given point in time for the hospital administrator the statistical model still needs to adjustment of the standard errors for clustering in centres and per patient. Such an analysis describes a series of cross-sectional snapshots over time and is only as good as the cross-section of patients available at that point in time. With regards to missing data – maybe multiple imputation methods could be used as a form of sensitivity analysis to inform how much missing data may have influenced the results.

I guess that the authors of this paper had more the view of a hospital administrator and modeled the means of hemoglobin at different points in time as a function of other prevalent factors in the dialysis population. A major problem of this paper is that there is not enough information given about the study design to appreciate whether this was a representative study or not. Ideally the authors go back to the Swiss registry and compare their patient features at baseline with the features of the Swiss dialysis patients overall. Probably they are right in that there is not too much selection bias, given that the baseline distribution of primary renal diagnoses and comorbidities resembles UK data – but this absolutely needs to be substantiated with above comparison before the authors can even address their study aims. At present it is impossible to reach any conclusions from the analyses presented as long as this major point has not been dealt with.
The authors do not provide information on whether this study underwent ethical review. It may be true (though I am not aware of it) that audits of treatment quality don't need individual consent. However, data should only be used if it is ethically acceptable. Monitoring treatment success should be done with sufficient quality standards to ensure that patients profit from the auditing done – particularly so if they never were consented for using their data.

The analyses have to be repeated without using the LOCF method. Averaging over the whole sample should be per time point and standard errors should account for clustering of values in patients and centres. Analyses should adjust for case mix, age and sex, and should show measures of uncertainty/sampling error, e.g. standard errors or confidence intervals (for example figure 2 A & B). As this is a prevalent sample, the vintage of patients on dialysis should be displayed, and its effect on Hb level assessed.

Most of the tables and figures can be condensed or put into tables (for example Fig 3-5). Some of these figures (for example Figures 3 and 4) may even be misleading as they probably show aggregate (all time points) crude data which are not adjusted for time-point of measurement, age, sex and case mix.

A minor point is to move definitions of iron status (adequate iron status, functional, absolute iron deficiency) from the results into the methods section.

**Level of interest:** An article of importance in its field

**Quality of written English:** Acceptable

**Statistical review:** Yes, and I have assessed the statistics in my report.

**Declaration of competing interests:**

I declare that I have no competing interests.