Reviewer’s report

Title: Predictors of mortality of patients newly diagnosed with clinical type 2 diabetes: a 5-year follow up study

Version: 3 Date: 14 April 2010

Reviewer: Jan Cederholm

Reviewer's report:

1. In my last review I was still puzzled by the presence of age mentioned as one of the predictors (covariates) in Table 1. This was the reason why I asked if mean survival time, as mentioned in the Statistical section, was based on the observed follow-up period for data in Table 2. However, the authors repeat in their reply that age was used as the time scale in the analyses, as also mentioned in the Statistical section. The authors have now also corrected to use life expectancy, instead of mean survival time, both in the Statistical section and in Table 2.

However, this raises a question: how can age be used as the time scale for the Cox in Table 1, if age also is presented as one of the predictors (covariates) in this Table 1?

As far as I can see, there are two main choices for the time scale in the Cox, time-on-study (=duration in this sample of newly diagnosed patients) or age at survival (the age at which a patient experiences a total mortality event or terminates participation in the study). If total mortality is the event, birth is the point of origin for this age, and it can be correct to use this age as the time scale for analysis of total mortality. Furthermore, Korn et al (ref 17 in the latest revised version) have argued that a model with age as time scale is generally more appropriate. This has e.g. compelled the European SCORE project to model age in estimation of 10-year risk of fatal CVD.

However, Pencina and D'Agostino at Boston University have underlined that the Framingham study has consistently modelled time-on-study as time scale in research concerning CHD and CVD. Furthermore in a paper 2007 (Statist Med 2007;26:1343-1359) they show with examples from Framingham, that use of time-on-study or age as time scale make no real difference, which they were able
to demonstrate by estimation of regression coefficients for some included covariates as risk factors for CHD, e.g. diabetes and smoker. However, they underline strongly that when using age as time scale, it is necessary to adjust for age at study entry. If using an unadjusted age scale, this will give incorrect estimates of the regression coefficients, e.g. the coefficient for diabetes was then found to be weak and non-significant in their modelled example, although diabetes is a strong risk factor for CHD. The unadjusted age model also had a poor calibration compared to age or time-to-study models adjusting for age at diagnosis.

Thus, as I can see the situation in this article: age at survival = age at diagnosis + diabetes duration. This means that the predictor age found in Table 1 is the same as age at diagnosis. This also means that the authors have correctly used age at diagnosis as an adjusting covariate, when using age as time scale for the Cox regression in Table 1.

This also explains that HR for age at diagnosis is less than 1 in Table 1, although mean age at diagnosis was higher in the deceased patients than among survivors in Table 1. I can see that the authors now have included more explaining text regarding this in the Statistical section, which is clarifying.

I suggest that the name age at diagnosis is used for this covariate also in Table 1, as the authors now have named this variable as age at diagnosis in the explaining text of the Statistical section. This would make it more easy for the readers to understand this variable in Table 1.

2. I find it acceptable that the authors now have included my suggested more recent reference for analysis of life expectancy (Nelson et al, 2007, ref 19 in this revised version). Although their previous reference included the most well-known statisticians Cox and Breslow, this reference was old, from 1972, and not easily found nowadays when searching references by the Internet. The term life expectancy is now used everywhere in the paper concerning the results in Table 2. The authors have preferred to calculate median life expectancy from the 0.5-point of the survival curves, instead of calculating life expectancy from the area under the curve.

Articles analysing life expectancy are not frequent in the literature, although e.g. a recent paper by Leal, Gray and Clarke from Oxford and Sydney have performed such analyses and presented tables of life expectancy related to various risk factors using the UKPDS Outcomes Model (Eur Heart J 2009;30:834-839). However, it should be noted that the UKPDS Outcomes Model is based on parametric proportional hazard risk equations. The reference by Nelson et al (19) has used Cox regression. It has also been suggested that the Gompertz distribution might be better for analysing total mortality, which is exponential (sigmoidal), and said to be suitable for analysing total mortality including patients of high age. However, SAS version 9 still has very little to offer regarding Gompertz, and I find it sufficient that Cox regression was used here for calculation of life expectancy.
3. Results of the test for the proportional hazard assumption of the Cox Model II is given in the text below in Table 1, that is adequate.

4. Control for interaction between self-rated health (SRH) and other covariates is important for reliable statistical results. This has been well performed. However, I suggest that the data for interaction in all patients should be included in Table 1, not in Table 6 as now.

5. I agree with the authors that confounding, and residual confounding, is a most important issue in epidemiology, especially in research about SRH, and this has been sufficiently addressed in this article, according to available baseline data, and with the now added complementary analysis in all patients of confounding by urinary glucose and several symptoms at diagnosis, and with a reasonable underlining by the authors that the possibility of residual confounding still cannot be ruled out.

6. The authors have now fully responded to my comment regarding the clinical usefulness of this model for SRH-rating, as only excellent health differed considerably from all other alternative ratings. I agree that the fully adjusted Model II should be used for conclusion on these findings, and the authors have now also added trend analysis on the SRH-rates.

However, I still say that the results imply that this model for SRH seems not too suitable for general clinical use. An alternative is the EQ VAS Score, using a thermometer-like scale from zero (worst) to 100 (best). Hayes, Clarke et al have demonstrated that a multivariate adjusted 10-point higher EQ VAS score was associated with a 6% (95% CI 1–11) lower risk of vascular events, and a 22% (95% CI 15–28; p<0.001) lower risk of diabetes complications in a larger study of 7,000 patients with type 2 diabetes (ref 6 in this paper). They used time-on-study as time scale at Cox regression, and adjusted also for diabetes duration. EQ VAS Score and this reference is only mentioned by the authors in the Introduction as "evidence of SRH is scarce for people with diabetes", and in the Discussion as "evidence of associations … SRH with morbidity and mortality … is only slowly emerging for people with type 2 diabetes".

I suggest as a minor revision that the authors should somewhat more discuss on the clinical usefulness of the SRH-rating in this paper, compared with the results using the EQ VAS Score in the study by Hayes, Clarke et al.

However, I find the results in Table 2 interesting, that this questionnaire was able to show a reduced life expectancy for all other ratings of SHR versus excellent SHR, quite comparable to the reduced life expectancy caused by a history of CVD which is a strong risk factor for mortality. The authors conclude that doctors should be more aware to discuss perceptions of health with patients having newly onset diabetes, and to be attentive of sub-optimal self-rating. This should add to clinical skills.

**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