Author's response to reviews

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

Authors:

Niels DF Olivarius (olivarius@sund.ku.dk)
Volkert Siersma (siersma@sund.ku.dk)
Anni BS Nielsen (annibrit@sund.ku.dk)
Lars J Hansen (larshan@sund.ku.dk)
Lotte Rosenvinge (lotte.rosenvinge@dadlnet.dk)
Carl Erik Mogensen (carl.erik.mogensen@ki.au.dk)

Version: 3 Date: 7 April 2010

Author's response to reviews: see over
To the editors of BMC Endocrine Disorders 7 April 2010

Submission of a revised version of the manuscript “Predictors of mortality of patients newly diagnosed with clinical type 2 diabetes: a 5-year follow up study”

Thank you for inviting us to resubmit our manuscript to BMC Endocrine Disorders.

The fourth reviewer made it clear that - despite our revisions after the first review - several aspects of the methods need to be explained in more detail, which we have done. We have also clarified these aspects of the methodology in the manuscript.

Below is our answer to the critique (in New Times Roman and blue) and our major text revisions (in italics).

Please find attached a Word-comparison of the former (old) version of the manuscript with the revised (new) version.

We are looking forward to your reply.

Sincerely yours,

Niels Olivarius

MS: 3327425223420145
Predictors of mortality of patients newly diagnosed with clinical type 2 diabetes: a 5-year follow up study
Niels DF Olivarius, Volkert Siersma, Anni BS Nielsen, Lars J Hansen, Lotte Rosenvinge and Carl Erik Mogensen

Dear Prof Olivarius,

Your manuscript has now been peer reviewed and the comments are accessible in PDF format from the link below. Do let us know if you have any problems opening the file.

Referee 1:
http://www.biomedcentral.com/imedia/2026388281358950_comment.pdf

Referee 2:
http://www.biomedcentral.com/imedia/5537820843619094_comment.pdf

We would be grateful if you could address the comments in a revised manuscript and provide a cover letter giving a point-by-point response to the concerns. Please note that this revision may be required to under-go re-review with one or both referees post-submission.

Please also highlight (with 'tracked changes'/coloured/underlines/highlighted text) all changes made when revising the manuscript to make it easier for the Editors to give you a prompt decision on your manuscript.

Please also ensure that your revised manuscript conforms to the journal style (http://www.biomedcentral.com/info/ifora/medicine_journals). It is important that your files are correctly formatted.

We look forward to receiving your revised manuscript by 25 March 2010. If you imagine that it will take longer to prepare please give us some estimate of when we can expect it.
You should upload your cover letter and revised manuscript through http://www.biomedcentral.com/manuscript/login/man.asp?txt_nav=man&txt_man_id=3327425223420145. You will find more detailed instructions at the base of this email.

Please don’t hesitate to contact me if you have any problems or questions regarding your manuscript.

With best wishes,

Mick

Mick Aulakh, M.Sc.
Assistant Editor, BMC Medicine
BMC-series
BioMed Central

Reviewer’s report
Title: Predictors of mortality of patients newly diagnosed with clinical type 2 diabetes: a 5-year follow up study
Version: 2 Date: 21 February 2010
Reviewer: Jan Cederholm

Reviewer’s report:
I find that the authors have responded sufficiently to my questions regarding major and minor revisions in the previous version 2, and have changed the manuscript accordingly.

I have only two minor revisions for this version 3.

1. Estimation of mean survival time has now been explained in the statistical section. However, Table 2 should use the same terminology: mean survival time, instead of life expectancy.

As far as I can understand this method, it estimates mean survival time based on the observed follow-up period.

Analysis of life expectancy is often carried out with a method for survival analysis, where age is substituted for time in the regression, in order to allow extrapolation beyond the observed follow-up period. Median life expectancy can be calculated as area under this predicted survival curve. See Nelson et al as reference (Statistics In Medicine 2008;27:5525-5555).

Thank you for this statistical reference, which we find so useful that we have replaced the Breslow-reference with this newer reference from 2008. As we have explained below, we are using age as time scale in these analyses which is why the correct term is life expectancy. We regret that we have used the term survival time in the former version of the manuscript. It has now been replaced by the term life expectancy. We have changed the Statistical analysis section to describe the choice of age as time scale, as described below. The statistical procedure for Table 2 was simple: First we calculated the predicted survival curves. Then the median life expectancy was calculated from the 0.5-point of these survival curves. We did not make use of area under the curve, which would have given the mean life expectancy.

The text in the Statistical analysis section has been changed accordingly:
The effects of selected characteristics were illustrated in Table 2 by the median life expectancy. This was projected by model II through the 0.5-point of the survival function as estimated by the method suggested by N. Breslow in his discussion of a paper of D. R. Cox [1, 2].

2. A main result of this study is that less than excellent self-rated health (SRH) is a risk factor for 5-year total mortality. However, how should a clinician act if comparing two patients with good or fair/very poor SRH, which also were the most frequently reported SRH alternatives? This should be somewhat more commented in the Results and Discussion sections.

We have changed the Results and Discussion to clarify this. See below (reviewer 4, point 1). We have been very cautious in the discussion of the clinical implications of our findings. As discussed 8 lines below we cannot discuss the implication of differences between the effects of the three non-excellent SRH-categories as we did not find any differences between the three effects in model II.

As far as I can see in Table 1, a weaker trend of increasing HR for risk total mortality, when comparing good or fair/very poor SHR with excellent SHR as reference, still remained when adjusting for risk factors according to Model 1 in Table 1, although this trend was less pronounced than at the univariate analysis.

We consider that Model II in Table 1 contains the main result of the analyses. In this model the HRs for the four categories of SRH are (1), 2.46, 2.51 and 2.16. No trend is perceptible, and the p-value from a Wald test for equality of the three SRH effects is 0.80. Even in Model I the trend is weak: (1), 2.60, 3.02, and 2.97 (p = 0.22, Wald test). A trend test is not incorporated in the standard statistical software in these analyses, but we have added a simple trend test to Table 1 (footnote f) and consider this test to be a supplement to the Wald test:

“/ Trend test including self-rated health as a continuous variable in the multivariate regression analysis: p-value < 0.0001 (model I); p-value = 0.091 (model II)”

and to Table 3 (footnote f):

“/ Trend test including self-rated health as a continuous variable in the multivariate regression analysis: p-value = 0.027 (model I); p-value = 0.55 (model II)”

and to Table 4 (footnote f):

“/ Trend test including self-rated health as a continuous variable in the multivariate regression analysis: p-value = 0.078 (model II)”

In both model I and II the difference between the level of HR for 1) excellent and 2) non-excellent categories is considerable, and a Wald test seems to be the most appropriate supplementary statistical test to document our main statement that less than excellent health is a risk factor for mortality. We have changed the Results section accordingly:

“[(new text as indicated below, see reviewer 4, concern no. 3)] SRH was still an important predictor, but in model II no longer with any perceptible trend (Table 1, footnote f), and the effects of the three non-excellent categories of SRH did not differ in both model I and II (Table 1, footnote e).”

Level of interest: An article of importance in its field
Quality of written English: Acceptable
Statistical review: No, the manuscript does not need to be seen by a statistician.
Declaration of competing interests: I declare that I have no competing interests

Reviewer’s report
Title: Predictors of mortality of patients newly diagnosed with clinical type 2 diabetes: a 5-year follow up study
Version: 2 Date: 1 March 2010
Reviewer: John Yudkin
Reviewer’s report:
This reviewer was asked to comment on the revised manuscript in order to see
whether the comments of Reviewers 1 and 3 had been adequately answered in the latest draft.

Olivarius et al have analyzed the predictors of mortality from a population-based sample of 1,323 persons newly diagnosed with clinical diabetes and age above 40 years. The main topic of the present manuscript was the evaluation of a simple questionnaire about the self-rated general health (SRH). SRH was evaluated with a single question: in general, how would you rate your health at present? The response categories were excellent, good, fair, poor and very poor. The authors concluded that patients who rated their health as less than excellent had increased 5-year mortality, even when biochemical, clinical and lifestyle variables were controlled for.

The authors have clearly put in an enormous amount of work to perform the analyses for this and previous drafts of the manuscript, which they clearly think is warranted by their conclusions. This reviewer feels that in general the authors have dealt with the anxieties of Reviewers 1 and 3, and these points will be considered below. There are, however, a number of separate problems, many of these being related to the methods and interpretations of the statistical analyses. As this reviewer has no statistical expertise, it is important that the comments of the previous reviewer 2 are re-requested, but there is an additional problem. This reviewer is probably around the median when considering the knowledge of the readers of BMC Endocrinology on statistical and epidemiological methods. This in turn means that things which confuse this reviewer may well confuse the audience, if the paper is accepted. Indeed I note that one of my main puzzles, the hazard ratios for age, were also not clear to the reviewer (2) with major statistical expertise. Thus much clearer explanation is called for. Additionally, while the majority of readers may be familiar with Information Criteria AIC and BIC and their significance, this is not something with which this reviewer is familiar. It might be useful to expand the explanation.

I will now itemize my concerns, starting with the major ones which have not been raised by other reviewers.

1. The main problem with the measure of self-rated health is likely to be confounding. The authors do not show the relationships of self-rated health with any of the variables collected,[…but we assume that these associations may exist. We do not report them because they are beyond the focus of the manuscript] and deal with the confounding issue by multivariate Cox analyses. The fact that inclusion of pre-existing disease substantially reduces the HRs, and the significance of the Wald test is reduced from P=0.007 to P=0.036, leaves open the possibility of residual confounding, bearing in mind that disease was self-reported. Furthermore the relationship of self-rated health with age is important [When the interaction between age and SRH was added to model II (Table 1) p was 0.38 (Wald test, Table 6, last column).], particularly bearing in mind that it is not clear to me whether chronological age is included in the analyses (see below and find our answer below too). Residual confounding is certainly an important issue in epidemiology, especially in research about SRH, and it is the main issue in this manuscript. We have already addressed the question of residual confounding in the manuscript, but we are now more explicit in the following addition to the Discussion section:

“Responders’ grounds for rating their health probably represent a personal estimation of longevity [3], which may be related to current or previous physical health, symptom perception, personal resources and physical functioning [4,5] together with health behaviour [4,5], comparison with the health of age peers [5], and a knowledge of familial dispositions [3]. Until now there is only weak evidence to suggest that mental states affect clinical outcomes independently of conventional risk factors [6]. In the present analyses a clear trend of increasing mortality with decreasing SRH was observed univariately (Table 1), but after multivariate adjustment with all available possible confounders the three non-excellent SRH categories were no longer different (Table 1, Model II, footnote e). This means that in these patients newly diagnosed with diabetes the effect of SRH boils down to whether their health is considered to be excellent or not. It seems reasonable to assume that SRH carries risk information which cannot fully be uncovered by the clinical information available at diabetes diagnosis. The possibility of residual confounding, however, cannot be ruled out and it is probable that these unknown confounders to a considerable degree have a biological basis [7,8].”
2. The authors note that the likely reason that the expected relationships of risk factors with mortality is not seen may be because of the metabolic disequilibrium within a week or two of diagnosis of diabetes. But this metabolic disequilibrium may well also impact on self-rated health. The authors have explored self-rated health in people who have just been diagnosed with diabetes, and who are in the majority of cases likely to have had symptoms related to hyperglycaemia and polyuria for a variable period of time. It seems likely that self-rated health would therefore be affected by the degree of elevation of plasma glucose level, the renal threshold and by the duration of symptoms. These last two are unavailable, but diagnostic plasma glucose is.

First of all we would like to stress, as we have done in the manuscript, that we consider it to be an important feature of these analyses is that the patients were newly diagnosed. The patients were examined by their GP on average 12 days after diabetes diagnosis, and we believe that it is interesting in itself to see which baseline factors are related to mortality in this very special situation in which the patient has to make important decisions about his or her future life. Our results, however, cannot be extrapolated to patients with known diabetes of different duration. The metabolic disequilibrium probably had an impact on SRH, but this does not invalidate our aim as indicated above. We have approached the reviewer’s concerns about renal threshold and diagnostic symptoms by making an extra analysis, which has been added to the Results section:

“As SRH may be related to renal threshold and presence of symptoms, we added urinary glucose at diagnosis (linearly, $p=0.78$) and number of 16 specified symptoms at diagnosis (0, 1, 2, 3, and 4 or more symptoms, $p=0.93$) [9] to model II (Table 1). This analysis gave the following hazard ratios (95%-confidence intervals) for good, fair and poor/very poor SRH, respectively: 2.42 (1.26;4.64), $p=0.0081$; 2.50 (1.32;4.73), $p=0.0050$; and 2.16 (1.04;4.49), $p=0.040$. I. e. the association between SRH and mortality was virtually unchanged.”

3. The consideration of age in Table 1 and 3 is analysed as follows: ‘death intensity was represented as a function of patient age, multiplicatively affected by the characteristics.’ It is not clear to me how this was done. If this was adjusted for within the sample, then the effect of age should have disappeared. But it now seems that younger patients had a ‘relatively higher mortality than older patients (reviewer 2 point 6). This is likely to be incomprehensible to the average reader.

Looking back, we might have chosen another time scale for these analyses, but the choice was made after careful consideration of the advantages and disadvantages of different time scales. We ended up deciding to make the multivariate Cox proportional-hazard regression analyses with the patient’s age as time scale. We adjusted for late entry into the risk set since the patients entered the risk set at age of diabetes diagnosis, not at birth. In conventional Cox regression analysis, where the time scale is time-in-study (or diabetes duration), a diagnosis-age effect would compare the mortality hazards of those diagnosed at a young age with those diagnosed at an older age for each time point after diagnosis. I. e. a hazard ratio of 2 would indicate that those diagnosed at an older age have twice the hazard of those diagnosed at a young age given that they have spent equal time in the study. This estimate naturally comprises the higher mortality of people of older age, and many readers probably expect to find this estimate in a paper. However, we chose to use age as time scale because we considered that the hazard changes more with age than with diabetes duration. Please find a discussion of the pros and cons of using age as time scale in this reference: [10]. In such analyses, a diagnosis-age effect compares the mortality hazards of those diagnosed at a young age with those diagnosed at an older age at each age; a hazard ratio of 2 now indicates that those diagnosed at an older age have twice the hazard of those diagnosed at a young age given that they have the same age (and hence different time-in-study/diabetes duration). Here the effect that people of high age have a higher mortality is included in the baseline hazard and the diagnosis-age effect shows that patients diagnosed at a young age have a more severe disease than those diagnosed at an older age (since we found that HR was less than 1), i.e. the age-specific mortality hazard is lower for those diagnosed old relative to those diagnosed young.

We have extended the first paragraph of Statistical analysis with some of these considerations. We will be happy to include further details if you find it necessary:

“The influence on all-cause mortality of baseline characteristics measured at diabetes diagnosis, i.e. socio-demographic, clinical, biochemical and behavioural variables as well as complications and SRH, was investigated in Cox proportional hazard regression models. In these models the death intensity was represented as a function of patient age, multiplicatively affected by the characteristics, and patients entered the risk set at the time of diagnosis. The effect of a
characteristic was assessed by a hazard ratio, i.e. the ratio of the age-specific mortality rate in a specific category of a patient characteristic compared to the mortality rate in a reference category. We chose to use age as time scale instead of diabetes duration, which is more common, because we considered that the hazard changes more with age than with diabetes duration [10]. In such analyses, a diagnosis-age effect compares the mortality hazards of those diagnosed at a young age versus those diagnosed at an older age at each age, and a hazard ratio of e.g. 0.5 indicates that those diagnosed at an older age have half the hazard of those diagnosed at a young age given that they have the same age (and hence different diabetes duration), i.e. the age-specific mortality hazard is lower for those diagnosed at an older age relative to those diagnosed at a young age. We used the PROC PHREG procedure from the SAS statistical package ver. 9.1.”

In Results we have changed a sentence in order to clarify this methodological challenge:

“In the three multivariate models in Table 1 the predictive value of urinary albumin, resting heart rate and smoking became non-significant. Relatively low age at diabetes diagnosis increased the relative risk of death, and so did Patients diagnosed at a young age had a higher age-specific mortality than patients diagnosed at an older age, and male gender, low body mass index and low level of physical activity also increased the relative risk of death.”

4. There are a number of variables in tables 1 and 3 where the HR from the Cox model differs greatly from an RR calculated from the total numbers. Thus the HR for living alone is 1.07 but I calculate the RR as 1.81. Conversely, 35% of survivors and 33% of deceased are current smokers, yet the HR is 1.71.

5. In the opposite direction, there are powerful influences of heart rate, despite the median being 76 in both groups.

4+5: The (approximate) relationship between a HR from a Cox model and a RR calculated from the total numbers holds in the traditional setup of the Cox model where the time scale is diabetes duration (as described above). The HR from a Cox model with age as time scale cannot be compared with the RR as calculated from the numbers given in the tables.

6. The measures of blood pressure clearly show that the 474 GPs in the study used digit preference. But this in turn raises the issue of inter-observer variability when the mean number of subjects is 3 per doctor.

A model like model II where a robust sandwich estimator for the covariance matrix is used to account for model misspecification (e.g. ignoring clustering), gave the following hazard ratios for good, fair and poor/very poor SRH, respectively: 2.46 (1.22;4.97), p=0.012; 2.51 (1.24;5.05), p=0.010; and 2.16 (0.99;4.71), p=0.053. This method of dealing with the challenge of clustering/inter-rater variability is now described in the Results section in the manuscript as follows:

“Additional analyses
The question of inter-rater variability was analysed as a case of clustering. A model like model II from Table 1, where a robust sandwich estimator for the covariance matrix was used to account for clustering, gave the following hazard ratios for good, fair and poor/very poor SRH, respectively: 2.46 (1.22;4.97), p=0.012; 2.51 (1.24;5.05), p=0.010; and 2.16 (0.99;4.71), p=0.053. i.e. we found no evidence that inter-observer variability has influenced the results [...]”

7. The issue of laboratory standardization was raised by Reviewer 1 but not answered.

Yes, I can see that we regrettably concentrated on the question of variability. All blood samples were determined at The Department of Clinical Biochemistry at Odense University Hospital in cooperation with the head of department, Professor Mogens Hørder who took responsibility for blood chemistry [11]. Quality assurance was obtained through routine procedures with commercial control preparations. This has been described for HbA1c in a separate paper [12]. We have added this sentence to the Assays section in the manuscript:

“Assays
[...] All remaining blood samples were analysed at Odense University Hospital. Serum total cholesterol was determined enzymatically with cholesterol esterase-cholesterol oxidase-peroxidase reagent and fasting serum triglycerides with a lipase-glycerokinase-glycerol-3-phosphate oxidase-peroxidase reagent. Urinary albumin concentration was measured in a freshly voided morning urine at Arhus University Hospital by a polyethylene glycol radioimmunoassay [13]. Quality assurance was obtained with commercial control preparations.”

8. The concept of ‘person-years deceased’ (Results line 6) is a curious one.
We were asked to provide this figure by reviewer 3. The figure has now been removed from the manuscript in the first paragraph of the Results section:

“The total number of person-years in the study was 6447 (survived: 5568, deceased: 880; men: 3084, women: 3363).”

9. Table 4 should include confidence intervals. These have been added to Table 4.

10. It is bizarre to include data from Table 1 as a column in Table 6. This pragmatic choice was made primarily because Table 1 is very large even as it was. To avoid confusion we have now moved the last column of Table 6 to the last column in Table 1.

I list below the queries of the Reviewers 1 and 3 which I feel were not satisfactorily answered.

Reviewer 1
(2) Any valid standardization of lab values?
Please find our answer above (concern no. 7 of reviewer 4).

(3) "In table 4 the highest 5 years CVD mortality was noted in patients with good SRH versus excellent SRH, but the HR was much lower for the category "poor". Date by chance to due to much smaller sample size? By contrast the predictors of all-cause mortality in patients who died within 3 years of diabetes diagnosis are more plausible."
The authors reply that ‘In the subgroup analysis in Table 5 (new 4) of “healthy” patients without any known chronic conditions or complications (from a specified list, see Table 5 (new 4)) at diabetes diagnosis, the trend still exists in the univariate analysis and disappears in the multivariate model’.
Is this a formal test for trend or are they merely eyeballing the numbers?
We admit that we thought that eyeballing sufficed in this case. We have now made a new Wald test and added it to Table 4 as a new footnote:

"* Wald test for the equality of the three self-rated health effects: p-value = 0.71 (model II)"

The same question applies on p10 line 1 with regard to Table 1. In this case we had already included two Wald tests in footnote e in Table 1, and we are now referring directly to this footnote in the manuscript as indicated above (concern no. 2 of reviewer 2(JC)):

"* Wald test for the equality of the three self-rated health effects: p-value = 0.22 (model I); p-value = 0.80 (model II)"

As indicated above, we have also included a supplementary trend test in the relevant tables and changed the Results section (see our answer to concern no. 2 of reviewer 2).

Reviewer 3
(1) “Although DM was newly diagnosed in 1323 patients at the baseline of the study, 298 (22.5%) patients of them died during a 5-year follow-up. Even in relatively healthy subjects without any other diseases (n=696, Table 5), the mortality was 14.5% (101/696).”
While the authors have responded to the issue of treatment, they do not tackle the question of the enormously high mortality. The 5 year mortality in the conventional treatment limb in ADVANCE (mean age 66) was 9.6%, 3.5 year mortality in ACCORD (mean age 62 with known CV disease) 4%, and 10 year mortality in UKPDS (mean age 53) 18.9%.
In e.g. UKPDS age at entry was 25-65 y. In the study ‘Diabetes care in general practice’ we included all persons with newly diagnosed diabetes and aged 40 y. or over. In our paper comparing the over-all mortality of our patients with the background population, the 5-year mortality was 16.8% (104/619) for women and 21.6% (152/704) for men [14]. The number of dead with an age at time of death above 64 years was 94 of 104 for women and 113 of 152 for men.
Furthermore our population-based sample cannot be compared directly with other patient samples. Quite correctly these facts may not be known to all readers, and we have added a sentence to the end of the first paragraph of the Discussion section:

“The high mortality rate in the present study must be viewed in light of the fact that patients were included with no upper age limit.”

(2) “A part of patients had cardiovascular disease (378 patients) or cancer (63 patients) at baseline of the study. Naturally, the two diseases influence baseline serum tests, BMI, heart rate, blood pressure, lifestyles and SRH. Subsequently the two diseases increase mortality of the patients, and 141 (51%) patients out of 278 patients deceased had cardiovascular disease at baseline (Table 1). In such a DM population with other severe diseases, predictors of mortality of DM patients can not correctly be drawn. Therefore, the patients with the two diseases at baseline should be excluded from the study.”

The issue of confounding, and residual confounding, is made in my point 1. We believe that we already have taken these concerns into consideration. Almost all of the analyses in the many tables have the purpose of falsifying the identified association between SRH and death. The possibility that the association may be explained by residual confounding certainly exists, but we are not aware of any paper about the SRH-mortality relation which has been so thorough in its analysis of known and unknown possible confounders. After all these multivariate and subgroup analyses, SRH still has information value. The answer which we gave to this question to reviewer 3 on 3 February 2010 is still valid:

We agree that CVD, cancer and other chronic conditions at diabetes diagnosis are very important confounders of the relationship between SRH and mortality and as such these variables are included in Model II (e.g. in Table 1 (new 1)). Moving from Model I (without chronic conditions and complications) to model II (with these variables), the effect of SRH is lessened, but it is still statistically significant. These models are made on the basis of specific hypotheses (see below) and they are methodologically sound. Undiagnosed conditions, however, may have contributed to precipitate the diabetes diagnosis, and these conditions may be associated with both poor SRH and high risk of death. We approached this unobserved heterogeneity among our patients in their susceptibility to dying in three additional analyses, one of which was an analysis of “healthy” patients (without CVD, cancer, diabetic retinopathy, peripheral neuropathy and albuminuria) in Table 5 (new 4). In this analysis the HRs were not changed substantially (compare Table 1 (new 1) with Table 5 (new 4)).

The question of confounding is also discussed above (concern no. 1 by reviewer 4).

(7) “Authors/scientists should estimate statistical models based on their hypothesis and previous evidence before the beginning of analysis. Post-analytic selection of the models using AIC/BIC may be just statistically logical, but may neither be always biologically logical nor be related with authors’ hypothesis. For an example, comparison of results from Model 1 and Model 2 in Table 1 suggests that SRH are associated with yes/no of cardiovascular disease and subsequently that OR by SRH and OR by yes/no cardiovascular disease are confounded with each other. Therefore, Model 2 may not be suitable for the study objectives.”

While this reviewer feels that Model 1 is a reasonable test of their hypothesis, and Model 2 attempts to correct for confounding, the views of a statistical reviewer would be important to judge the merit of the authors’ response. The overall purpose of the presentation of the models is exactly as indicated by reviewer 4. Akaike’s Information Criterion (AIC) and Bayes’ Information Criterion (BIC) are two types of scores that assess the quality of the model relative to both fit and size. Naturally, when variables are added to the model, even irrelevant ones, the model fit (as measured by e.g. the likelihood) improves. The improvement of the model fit then has to be seen relative to the increase in size. This is what AIC and BIC do: they perform the trade-off (each in a slightly different way) between the model likelihood and the number of parameters in the model. The lowest value of the information criterion gives the model of best quality. The information criteria are used for quick comparison of models in a model portfolio (that do not have to be nested). For nested models (e.g. model I is nested in model II) comparing the AIC is similar to doing a Likelihood Ratio (LR) test. We have tried to clarify their use by introducing some of these considerations in the Statistical analysis section:
“Akaike’s Information Criterion (AIC) and Bayes’ Information Criterion (BIC) were used to compare the quality of the models relative to both fit and size [15]. When variables are added to a model the model fit improves, and this improvement has to be seen relative to the increase in model size. The scores of AIC and BIC perform the trade-off, each in a slightly different way, between the model likelihood and the number of parameters in the model. The lowest value of the information criterion indicates the model of best quality. Overall comparison of the models was done by investigation of the information criteria (AIC or BIC), i.e. the likelihood score penalised by the number of parameters in the model [15] where the lowest score indicates the best model.”

(8) “Generally, statistical power for detecting interactions (p.8) between two variables is low. The sample size (n=1323) is too small correctly to test interactive effects.”

The authors would be able to increase their numbers of events by looking at 10y mortality. This might also allow them to look at cause-specific mortality, and in a way which obviates their very crude lumping together of different disease processes in something such as ‘CVD mortality.’

In Denmark the quality of the National Death Register has been affected by administrative changes during the past decade. If we were to include more deaths in this analysis we would have to go through hundreds of death certificates manually. Unfortunately, we do not have the resources for such new data retrieval.

Reference List


