Author's response to reviews

Title: Life style and longevity among initially healthy middle-aged men: prospective cohort study

Authors:

Trond Heir (trond.heir@medisin.uio.no)
Jan Erikssen (j.e.erikssen@medisin.uio.no)
Leiv Sandvik (UXLEDV@ous-hf.no)

Version: 5 Date: 21 June 2013

Author's response to reviews: see over
Referee 1:
Version: 3 Date: 8 February 2013
Reviewer: Jamie Seabrook
Reviewer's report:
The authors have done a very good job addressing all of my initial suggestions and concerns. The paper is well written, and I recommend the paper for publication.

The only one suggestion that I have is as follows: in Table 1, the authors report p values for baseline clinical and laboratory variables between non-smokers and smokers, respectively. Presumably, an independent samples t-test was used to compare mean differences between the two groups, whereas a chi-square test was used for the BMI categories. The statistical tests used for Table 1 should be included in the Statistical Analysis sub-section of the Methods.

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.

We agree. These statistical tests are included in the Statistical Analysis sub-section.

Referee 2:
Version: 3 Date: 18 February 2013
Reviewer: Aage Tverdal
Reviewer's report:
Life style and longevity among initially healthy middle-aged men: a prospective cohort study
I am not convinced that the method that required the exclusion of 41% of the study population gives valid estimates. I think the journal should ask an impartial statistician about the design of the paper. The authors refer to an article in JAMA which did not exclude any participants due to fact of not reaching 85 years within the follow-up time of the study. In that study all participants 45-68 years, except for baseline exclusions, were followed for 40 years.

Reviewer 2 and 3 differ in their comments about the statistical analysis presented in our manuscript. When they differ, we have chosen to follow the advices from referee 3. According to the aim of the study, it was not appropriate to include individuals that could not reach the age of 85 during the observation period.

The authors claim in their response that the findings in their study are valid for healthy men between 51 and 59 years. People who were 51 years when they were examined in 1975 or 1974 must have been excluded as they could not have reached 85 years by the end of follow-up.

According to reviewer 2 people who were 51 years when they were examined in 1975 or 1974 could not have reached 85 years by the end of follow up. However, in the method section we have stated that the inclusion period started in August 1972.
Thus it is possible to reach the age of 85 for a participant that was 51 years at inclusion.

The author’s statement about the lack of significant impact of cholesterol and systolic blood pressure as possibly being due to low number of participants saps my confidence to the study. If the study has insufficient power to detect an influence of cholesterol and systolic blood pressure, what has it sufficient power to detect? Perhaps a paragraph on study size should be included.

We have clearly stated that the insignificant results regarding cholesterol and systolic blood pressure may be due to the relatively low sample size.

Regarding table 4, I understand that I have misinterpreted the coefficients. Sorry! The inclusion of calendar year in the model: OK. I am not sure the two added paragraphs on the check of the model assumptions make it more readable. The first part concerning spotting of multicollinearity is OK. The second part referring to Pearson and Deviance residuals as well as DIFBeta and the Hat matrix diagonal do not tell the reader much. I am not sure what is meant by Deviance residual and Pearson residual. Pearson deviance is well known. It is referred to the book by Hosmer and Lemeshow (2. edition). I cannot find VIF (variance inflation factor) and DIFBeta in the book (at least in the index).

We consider the methods used to check the model assumptions to be appropriate. Further, referee 3 seems to have no objections to these methods.

The part on interaction between overweight and physical activity among non-smokers had been tested and found not statistically significant. This is no surprise. As the study lacks power to detect main effects of cholesterol and systolic blood pressure, it is even harder (requires larger study size) to detect interactions. What was the power to detect an interaction between overweight and physical fitness? The authors refrained from a study on never smokers as only 208 persons were never smokers. I refer to the abovementioned points on the power of the study.

We agree with referee 3 that post hoc calculation of sample size should be avoided.

Level of interest: An article whose findings are important to those with closely related research interests
Quality of written English: Not suitable for publication unless extensively edited
Statistical review: Yes, and I have assessed the statistics in my report.
Declaration of competing interests: I declare that I have no competing interests

Referee 3:
Version: 3 Date: 21 May 2013
Reviewer: Giorgio Bedogni
Reviewer's report:
GENERAL COMMENT
I was asked to review this paper as BMC statistical referee.
I read version 3 of the MS (05 Feb 2013) and the replies of the Authors to the Reviewers.
I was specifically asked:
1) whether the employed analytic strategy is appropriate;
2) whether sample size is enough to test the study hypothesis.
I will start by answering the two questions that I was specifically asked:
1. The choice of the analytic strategy, i.e. logistic regression, is reasonable in the context of the aim of the study, that is to establish whether selected midlife variables can predict that a person reaches 85 years of age. Of course, the price to pay is a loss of power compared to survival analysis where more subjects would be counted for power, not only those experiencing the outcome. However, the choice of the method should be dictated first of all by the study question.
2. If the Authors have calculated sample size BEFORE performing the study they should tell so and how they did it. I expect that a sample size was calculated for the original aim of the study but this may have not be done for the present analysis. A post-hoc calculation of sample size should be avoided because it is meaningless, eg.
http://ndt.oxfordjournals.org/content/early/2010/01/12/ndt.gfp732.full. 95% confidence intervals will tell about the precision of the estimates in every case. Whether the (possible) lack of calculation of sample size is a limitation should also be judged in the context of the available literature, something which I have not the necessary knowledge to do.

We are grateful for the support for our choice of analytic strategy.

Sample size was not calculated before the start of the original study (in 1972). We agree that that a post hoc calculation of sample size should be avoided.

I have also the following comments:
MAJOR COMMENTS
Please report 95% confidence intervals (CI) for all effect sizes, including frequencies. For instance, 37% of non-smokers vs. 23% of non-smokers survived up to 85 years. What are the corresponding 95% CI? They are important to convey the precision of the estimate.

In the revised manuscript we have included confidence intervals as suggested (tables 2 and 3).

MINOR COMMENTS
Please, describe in greater detail the univariable and multivariable regression models under Statistical Analysis. All predictors and their codification for analysis should be reported here. For instance, from Table 4, I infer that cholesterol and systolic blood pressure were modeled as continuous as their effect is coded as 1 SD increase. This should be mentioned under statistical analysis. Speaking of continuous predictors, did you check whether their uni- and multi-variable logits were linear? Or did you transform them appropriately when needed? How?
We have described in greater detail the univariable and multivariable regression models under Statistical Analyses as suggested. Further we confirmed that the univariate and multivariable logits of the continuous predictors were linear. This finding is included in the Statistical Analyses section.

Smoking is modeled in 3 different ways: yes vs no (table 1), 0 vs. 1-9 vs. >=10 cigarettes (table 2) and >=10 vs <10 (table 4). Table 1 is probably for descriptive purposes of the two groups and there is no reasonable alternative. Why did you switch from [0 vs. 1-9 vs. >=10 cigarettes] to [>=10 vs <10]? Was this because there was no statistically significant difference between the 0 and 1-9 groups in Table 2? Which test was used to assess this difference? I wonder what would happen by modeling the number of cigarettes as continuous. This will make the analysis more difficult but will increase power and make the estimate of the effect of the increasing number of cigarettes more solid. However, the most important point to me is that there is no way to tell in advance whether an n-tomization suggested by univariable analysis will hold at multivariable analysis.

In table 4 we split data in non-smokers and smokers. For smokers, we only have two categories ≥10 vs. <1-9. Unfortunately the category 1-9 was denoted “<10”. This is corrected in the revised manuscript. When smoking habits were recorded at baseline, number of cigarettes was registered in categories. Thus it is not possible to model number of cigarettes as a continuous variable. The n-tomization of the smoking data was determined by the categories used at baseline.

Why was age split at 55 years? Was this the median age of the sample at baseline? Wouldn't it be better to model age as continuous, at least in a preliminary analysis? This may give a better insight into the relations of interest even if one has no reason to suspect that the difference of 8 years between those age 59 and 51 years could impact on the probability of reaching 85 years.

The median age of the sample at baseline was 55 years. We agree that it is better to model age as continuous, and have done so in the revised manuscript.

Fitness was operationally defined as work divided per unit of weight and corrected for age by linear regression. Is this the standard practice in the field? Again, I would not risk imposing structures to the data with n-tomization (tertiles) UNLESS they are biologically or clinically sound.

Our definition of fitness as work divided by body weight is in accordance with the literature. We corrected for age by linear regression because this was suggested by referee when we published our first fitness paper in New England Journal of Medicine in 1993. We have used this definition of physical fitness in our later publications. As regards presenting fitness data in tertiles or quartiles, this is quite common in epidemiological publications focusing on the relation between physical fitness and mortality.

From this perspective, BMI has no problem in being treated as discrete (besides the obvious loss of power as compared to when it is modeled as continuous). How many underweight (e.g. BMI < 18.5 kg/m²) there were among smokers and not smokers?
Only four participants had BMI less than 18.5 kg/m2, two smokers and two non-smokers.

You performed a separate analysis for smokers and non-smokers. A way to (potentially) increase power is using smoke as predictor (in your case would be <10 vs >=10 cigarettes) in the same logistic regression model. This has added benefit of making possible to test X*smoke interactions (within the power allowed by sample size, of course). Have you considered this possibility? Even more power could be gained by testing continuous X*smoke interactions. I do not think that this way of approaching the problem is at odds with the suggestion you give in the Discussion to model separately smokers and non-smokers. This way one could compare the estimates presently made separately for smokers and non-smokers.

We have considered the possibility to analyze smokers and non-smokers in the same logistic regression model, with and interaction term for smoking. We have chosen not to do so because of recommendations in the epidemiological literature, and because we wanted to be able to compare our results with similar studies.

Please, note that the effect attributable to fitness (high vs low) in non-smokers is highly variable: from very small (1.11) to very high (3.08). I believe that this has to be discussed in the MS by saying that further studies with larger number of subjects reaching 85 years (from were power comes from) are needed to provide more precise estimates of this effect. Although not statistically significant, the effect size attributable to fitness in smokers is not very different (0.95 to 3.39). One wonders whether the uneven distribution of outcomes/subjects in the two groups may be partly responsible for this difference and whether the strategy of using smoke as covariate (see above) could give better insight into this finding.

We agree that the wide confidence intervals for the association between fitness and longevity illustrate the relatively low test power in our study. Thus, further studies with larger number of subjects reaching 85 years are needed to provide more precise estimates of this association. This comment is included in the revised manuscript.

Level of interest: An article whose findings are important to those with closely related research interests
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