Reviewer's report


Version: 1 Date: 8 July 2012

Reviewer: Susanna Toivanen

Reviewer's report:

Manuscript title


The manuscript by Christina Halford et al. investigates the associations between self-rated health (SRH) and use of social insurance and health care services (sick leave, disability pension, hospital admissions), and mortality. Questionnaire data from 8 population-based cohorts in Sweden linked to national routine registers is used. Baseline 1973-2003, with an average follow-up time of 10.5 yrs. (range 0-24.9 yrs.). Analyses are based on a study population of 11,880 men and women, age 25-99 years (14,470 observations). Effects of SRH at baseline on outcomes during follow-up are studied, and also for how long the effects of SRH persist during follow-up.

Main results

• In women and men, a significant negative association, adjusted for covariates, is found between SRH and sick leave, disability pension, and mortality, respectively.
• In men, a significant negative association, adjusted for covariates, is found between SRH and hospital admission.
• Associations between SRH and mortality found to be robust during the follow-up period.

Authors’ conclusions

SRH has a strong predictive validity in relation to use of social insurance facilities and health care services, and to mortality.

Reviewer’s comments

General comments

This is an important research topic. SRH plays an important role in public health monitoring as well as in research, and we urgently need proof of its predictive value for relevant health outcomes. The present manuscript establishes the relation between SRH and several important outcomes, not only mortality, but
also sickness absence, disability pension and hospital discharge. However, the results only confirm earlier findings and do not provide any new results. But the results are based on large population-based data sets, which is one of the manuscript's main strengths. Covering several important outcomes simultaneously perhaps makes the manuscript somewhat superficial. Nevertheless, there are a few problems, such as the somewhat crude handling of data and/or superficial description of research steps which makes it difficult for the reader to follow the reasoning of the authors. However, this gives room for improvement and several recommendations are given below to strengthen the manuscript.

First, short comments to questions suggested by the journal for reviewers:

1. Is the question posed by the authors well defined?
   Yes, the aim of the study is clearly stated, but is not consistent throughout the manuscript. For instance, some of the table headings are not entirely in line with the aims of the study.

2. Are the methods appropriate and well described?
   No, the methods are not sufficiently described. In fact, it is quite difficult for the reader to follow what is done in the study as regards to statistical analyses. Please make effort to improve this.

3. Are the data sound?
   Yes, data is very well suited for the aim of the study. However, description of how data is handled could be improved.

4. Does the manuscript adhere to the relevant standards for reporting and data deposition?
   Not really. Please check out a standard text for reporting epidemiological follow-up studies, such as the STROBE http://www.strobe-statement.org/

5. Are the discussion and conclusions well balanced and adequately supported by the data?
   Yes.

6. Are limitations of the work clearly stated?
   No, this could be improved.

7. Do the authors clearly acknowledge any work upon which they are building, both published and unpublished?
   Mostly, as far as I understand. Some additional references from Scandinavia are recommended.

8. Do the title and abstract accurately convey what has been found?
   Yes.
9. Is the writing acceptable?

The manuscript would improve if proof read by a native English speaker. Also, description of the methods and results sections could be improved a lot. Presently, it is difficult for the reader to follow the reasoning of the authors. It might me that things that are clear for the authors are taken for granted. This is the case in many scientific articles. To increase the readability and pedagogical value of manuscripts is our obligation as researchers. There are many ways the authors could proceed and some suggestions are presented below.

Major Compulsory Revisions (recommendations which the authors must respond to before a decision on publication can be reached)

1. Page 4, it is stated that “3,590 subjects were part of more than one subpopulation”. This means that one subject may be included several times in the analyses. How many subjects actually are included more than once? Could this be a problem? Could this affect the results?

2. Page 4, the use of the term/concept “occupational status” needs more explanation, justification, or at least a few valid references for the readers’ guidance as this is not a standard way of classifying occupational position, even less so of occupational status, from a theoretical (sociological) point of view, which has implications for the way of operationalizing occupational position (social class based on occupation). For more information, see e.g. Bihagen, E and Nermo, M. Social stratifiering och social klass. In Rostila, M and Toivanen, S (Eds.) Den orättvisa hälsan – om socioekonomiska skillnader i hälsa och livslängd. Stockholm: Liber, 2012, pp. 28-45.

3. Page 4, regarding the operationalization of occupational status, was the variable constructed from several items, or only from one question? How was the questions stated?

4. Page 5, regarding SRH. It is stated that SRH is measured with a well-established instrument using a seven-point Likert scale. In the analyses a three category SRH is used. How are these categories constructed and why? There is no information what so ever about the operationalizing of SRH which is quite alarming considering that the main focus of the manuscript is in fact SRH. This is not so convincing. Please clarify.

5. Page 6. The section “Ethical consideration” would be better placed on page 4 after the section of “Study population”. It is not in a relevant place now, between the “Data collection” and “Statistical considerations”.

6. Page 7. The authors state that “Missing data were not replaced”. Were subjects with missing data removed from the analyses? It is unclear on how many observations the linear and cox regressions were based on. This could preferably be stated in the Table headings or in the Tables somewhere. The reader needs this information in order to see whether the number of observations (=n) varies in the models and therefore make the models incomparable.

7. Page 7 and Table 3, regarding linear regression analyses.

a. The text that introduces the linear regression of number of days of sick leave
on SRH, adjusted for covariates, do not refer to Table 3, and thus the reader do not know where to look for the results presented in a table. Please include Table 3 in parenthesis in an appropriate place in the text on page 7 as well.

b. It is unclear in the text and in Table 3 which reference categories are used for the categorical independent variables and how the results should be interpreted. This needs to be explained for the reader. (The reader understands that the independent variables are categorical from the way the variables are presented earlier in the methods section).

c. The outcome variable is “number of days on sick leave” and one the independent variables, ”occupational status”, includes a category labeled “on sick leave or disability pension”. Does this mean that sick leave is captured both in the outcome (number of days on sick leave) as well as in the exposure variable (being on sick leave)? What kind of problems/bias could this maneuver introduce?

d. A modeling strategy for the linear regression is missing. Such a strategy would explain if there is a theoretical base for the order in which the covariates are entered in the analysis, or why are all the covariates thrown in simultaneously? Is the aim to look at the association between SRH and number of days on sick leave? Or is the aim to establish the “determinants of sick leave in multivariate analysis” as stated in the heading of Table 3? Then any of the covariates is as important as SRH? Well, this is not what the aim of the study stated. This should be clarified and adjusted accordingly.

e. What does “backward elimination of non-significant variables” mean? Which variables were eliminated? Did the statistical program decide which variables to enter or is there a theoretically based modeling strategy somewhere?

f. Is number of days on sick leave relevant, or is the sick leave spell more relevant, i.e. whether short or long term sick leave? Previous studies have tended to focus on long term sick leave as this form of sick leave is most harmful for the individual’s working life and health as well as for the workplace and the society as a whole. Could the analyses be run for short versus long term sick leave separately? SHR might have a different predictive value for long term sick leave?

g. As the association between SRH and number of days on sick leave was investigated in terms of linear regression then the time dimension of the data was not taken into account. Could this be done using a different statistical method of analysis? As it is now, the analyses are interpreted as cross-sectional associations? Are there possibilities and potential for more refined and sophisticated data analyses which are not taken into account in the present manuscript?

8. Page 7, paragraph 4, and Table 4, regarding Cox regression analyses.

a. A modeling strategy for the Cox regressions are missing, see point 7d above. Here it is unclear on what bases the covariates were chosen for the three outcomes: disability pension, hospital discharge, and mortality.

b. “with the outcome entered as the dependent variable”. ? This sentence should
be omitted or rewritten as the dependent variable is the outcome (and the independent variable is the same as covariate or exposure).

c. “and being on sick leave or disability pension (regarding the outcomes hospital admission and mortality)”. Does this sentence refer to the covariate “occupational status” described earlier in the methods section? Why is it now labeled using one of its’ categories? See also point 7c above regarding inclusion of this variable as a covariate.

d. Table 4, regarding the heading. See point 7d. What is the aim of the analysis? Establish an association between SRH and the three outcomes, adjusted for covariates, or to investigate a set of determinants for the outcomes, including SRH.

e. Table 4, the heading needs information on number of observations included in the models, years of baseline and follow-up.

f. Table 4, in addition to confidence intervals, is it necessary to report the p-value and chi2, as these are not commented in the results? P-value is explained in the text already.

9. Page 8, Results.

a. Paragraph 3. Regarding amount of days of sick leave per year among women and men, this number is less interesting when we do not have information whether this was a long term sick leave spell or several short spells during a year. This makes big difference. Again, could the analyses be conducted separately for long resp short spells of sick leave?

b. Paragraph 4. Regarding the results of SRH and sick leave, authors should explain how the results from linear regression should be expressed, what does it mean? One unit decrease of SRH increases the probability of sick leave by how much? It is important to explain for the reader how the regression analyses are interpreted, this increases the pedagogical value of the manuscript. As it is now, there is very limited information and the reader is asked to look at Figure 1. And all of a sudden the SRH is a categorical variable instead of a continuous as was described earlier in the methods section. No information is given on why and how SR was categorized. Lots of information is lost by categorizing this 7-point Likert scale variable. It would be much more interesting to plot the continuous SRH in Figure 1.

10. Page 9, Results.

a. Paragraph 1. Start a new paragraph after line 4, sentence that ends with “…, and 121 days.” when the analyses go from SRH and sick leave to deal with SRH and hospital admission. This is another outcome and another analyses and should not be mixed as it is now, it is confusing for the reader.

b. Paragraph 2. Authors state “Self-rated health was also significantly associated with being granted a disability pension after adjustment for the influence of the covariates.” Does this mean that a weak SRF increased the risk of disability pension or what? Or that a strong SRH decreased the risk of disability pension, and how much? Please explain properly so the reader can understand and follow.
c. Paragraph 3. Same questions as in point above 10b.

11. Page 16
a. Authors state for Figure 2 “Cumulative first hospital admission rate among women and men in groups according to self-rated health”. Here it is important to indicate that is the level of SRH that defines the groups. This applies for all the headings. Ant the grouping of SRH needs to be explained in the methods section as well. Why and how? Why three levels? Why not four? Why not use the SRH in continuous form?

12. Page 19, Table 3.
a. See point 7b above. Table 3 is not easy to interpret, the reader needs more information on reference categories, how to interpret the estimates and the results from the linear regression analyses.

13. Page 20, Table 4.
a. See point 8d above. How many observations are included in the models? Do we need all these statistics for significance in the table?

14. Pages 25 and 26, regarding Figures 5 and 6. The rule of thumb is that figures should display very important findings. If not, thy can be added in an appendix. There are 4 tables and 6 figures in this relatively short manuscript which is somewhat unbalanced. The authors are recommended to “kill their darlings” and choose those tables and figures that are essential for making a clear story of their results. Everything else can be omitted or maybe added in an appendix.

Minor Essential Revisions (recommendations which the authors can be trusted to correct)


2. Page 17, Table 1. The table is squeezed and difficult to read, could it be placed horizontally instead?

Discretionary Revisions (recommendations which the authors can choose to ignore)

1. One easy way of improving the readability of a paper is to keep the same order of the main terms (e.g. women and men, or sickness absence, disability pensions, hospital discharge and mortality) through the whole manuscript text as well as in the figures and tables. In the preset manuscript there is some jumping back and forth in the use of the main terms/concepts. For instance, look at the order of terms in the title of the manuscript and then in Table 4.

2. Relevant references of self-rated health are missing in terms of the work by Manderbacka and Lundberg:
a. Lundberg, O and Manderbacka, K. Assessing reliability of a measure of

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

**Quality of written English:** Needs some language corrections before being published

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

**Declaration of competing interests:**

I declare that I have no competing interests.