Author's response to reviews

Title: Multimorbidity Patterns and Health Service Use in Swedish 85 year olds: An Exploratory Study

Authors:

- Huan-Ji Dong (huanji.dong@liu.se)
- Ewa Wressle (ewa.wressle@lio.se)
- Jan Marcusson (jan.marcusson@liu.se)

Version: 3 Date: 29 May 2013

Author's response to reviews: see over
Response to

Prof. Dr. Hendrik van den Bussche & Ingmar Schäfer, M.A

Thank you for all your advice. We have made every effort to improve our manuscript. Below are the point-to-point responses to the queries. We hope that the responses have adequately addressed your concerns.

The response was done by all the authors together with statistician Ass Prof. Mats Fredrikson (Linköping Academic Research Centre).

Major compulsory revisions

Introduction

P4: Not every reader is familiar with the Swedish system. Please describe characteristics of access, organization and utilization of primary care and secondary care (transfer part of your discussion chapter here). Describe it in a way that explains the utilization data in table 1 become understandable (minimal contact frequency with GPs, low use of ER and hospital; is this related to “excess health” among Swedish octogenarians?)

Response: We agree. In this new version we have modified based on your advice and added the information about Swedish healthcare system in a fairly detailed extension in the end of introduction. The relative low GP visit (minimal contact) is possibly due to contacts with other caregivers (e.g. nurse contacts). In our previous study we reported that only 4% 85 year olds did not use any official health service (primary care, specialist care and ER).

Methods and Materials

P5: You mention that your research team reviewed the patients case reports. Please include information which data source provided these case reports. Are case reports identical to patient records? If yes, what morbidity (inpatient? outpatient? both?) do these reports cover and who keeps them? As for included morbidity, please describe precisely which diseases/diagnoses (single or grouped) were extracted when you used a closed list. If not, describe how you handled the probable multitude of data on diseases. P 5: Please give information about all measuring instruments you use, including those that measure the education degree.

Response: In this new version, we have added the information how we collected and interpret data. We have respond to the questions raised in the MM section. Concerning the level of education, we measured according to the years of education and used a cut-off of 7 years (also used in other aging study). Motivation is to minimize risk of collinearity with working status.

P.6: The results from cluster analysis are very sensitive against the options you choose regarding similarity measure, linkage procedure and stopping rules. Please give more information about the rationale for your decisions. 

- You state that you used the same cut-off distance in men’s and women’s cluster dendograms. Why did you choose the cut-off value of 15 and not any other value (e.g. 10 or 20)?
• You state in Discussion/Limitations (p.10) that there were differences in morbidity between men and women. What was the rationale behind choosing the same cut off values for cluster dendograms in men and women?

• Did you consider using stopping indices, e.g. Calinski–Harabasz pseudo-F index or Duda–Hart Je(2)/Je(1) index? Why did you decide against using a stopping index?

• Please give information why you used Yule’s Q measure instead of other measurements like the Jaccard coefficient, which would be suited for asymmetrical morbidity data?

• P6: Please define “cluster interaction”.

Response: We appreciate your advice and discussion about methodology. In the modified manuscript, we now report the coefficients (calculated based on Yule’s Q combined with average linkage between groups), representing the similarity between clusters and the rationale of cut-off. We also report the odds ratios associated with these coefficients in the terminal node of each cluster so that readers may have a more initiative understanding. Moreover, we compare our clusters with other studies so as to strengthen the rationale of this five-cluster solution.

• We have added Yule’s values instead of the SPSS rescaled cut-off value in figure 1 and 2. We hope the coefficients will improve the understanding rather than a visible inspection.

• Our preliminary ambition is to try to compare men and women’s cluster structures on the same similarity level of clusters. After having a look at coefficients of each clusters (or variables), we found that a cut-off between 0.2-0.3 were reasonable to reflect multimorbidity patterns. We chose five-cluster solution because neither smaller (six-cluster structure for women) nor larger (three- or four-cluster for men and four-cluster structure for women) cluster structures are privileged, because 1) given a cut-off of higher coefficient (smaller clusters), the structure would be one dominate big cluster (Cluster 1) together with all other small clusters. This solution would affect the next analysis (regression) by unequal predictors; 2) given a cut-off of lower coefficient, it could hardly be explained by some pathophysiological mechanism (combining men’s Cluster 3 and Cluster 4 or women’s Cluster 1 and Cluster 2 joining together); 3) a five-cluster solution identified clusters which were also reported in other studies (vascular cluster, cardiopulmonary cluster, etc).

• It would have been much helpful with stopping index. To our knowledge, both Calinski–Harabasz pseudo-F index and Duda–Hart Je(2)/Je(1) index are used for variables measured on a continuous scale[1] (page 127); Goodman and Kruskal’s gamma statistic are used for category variables (in this study, all morbidities were coded as binary variables) [1] (page 128). Yule’s Q is a special type of Goodman and Kruskal’s gamma (2x2 table). After going through the coefficients calculated by Yule’s Q in combined with average linkage between groups, we found: 1) in men’s clusters (figure 1), Yule’s Q value decreased from 0.317 to 0.231, so we decided cluster 3 and 4 should not be together (also from a clinical perspective, see above); 2) in women’s clusters (figure 2), Yule’s Q value decreased from 0.393 to 0.244 and 0.205. Statistically, cluster 3 should be divided into 2 small clusters (cut-off between 0.393 and 0.244), but it seems to be closed to reality (pulmonary and heart conditions) that cluster 3 should be considered as a whole cluster. It is also similar to the structure in another aging study [2].

• It is a good question whether co-absence make sense in the cluster structure and further decision. Jaccard coefficient and Yule’s Q are two of numerous binary similarity measures.
We chose this method (hierarchical cluster analysis with Yule’s value and average linkage between groups) based on earlier studies in this research area [2-4]. Yet, some other researchers tested Jaccard coefficient [5]. To our best knowledge, Jaccard coefficient \((a/a+b+c)\) does not concern negative matches [1, 5, 6]. Using odds ratios, Yule’s Q provides an intuitive explanation for binary variables (the similarity and association). In this sample population, not only the individuals with comorbidity but those who had 0-1 morbidity as well were included in the analysis. Since our purpose is to use the clusters to predict health service use, the information about negative matches (co-absence) could be useful if they can be concerned in the cluster decision.

Additionally, we tested Jaccard coefficient in the current sample (please see the resized figures here). Some pairs/clusters were similar and also closed to clinical real-life. But it seemed of great difficulty to make a clear cut-off decision (clusters with single morbidity), so it was limited to the latter approach.

In summary, no particular clustering method is best and should be definitely recommended. It has to be recognized that hierarchical clustering methods may give very different results on the same data [1][page 47 and elsewhere in the book]. Cluster analysis as a statistical method help us chasing after interpretable results in relation to the findings from clinical daily routines. We mentioned our opinions in the discussion part (see ‘methodological issue’).

- ‘Cluster interaction’ means one cluster’s effect affected by another one (the clusters’ effect is dependent of each other). We have added that in the statistical method in this new version. The purpose was to test whether or not multiple problems (a combination of clusters) would affect the models.

**Results**

P 6/7: You state that the response rate was 90%, because 586 of 650 individuals responded to your letter, but there were only 496 of 650 individuals that completed the questionnaire. Please consider this in calculating the response rate (i.e., please use the correct value of 76%). P6/7: Data on the distribution of the number of
diseases (degree of MM) are lacking whereas on p. 10 you speak about a “high prevalence of MM”

Response: We have modified our description in the current manuscript.

P.7: Please use ‘working status’ instead of ‘SES’ (see above). P7: Explain your definition of aging consequences. Did you proof this empirically? P7/8 and Tables 2/3/4/5:
• It is mandatory to publish all the variables included in the model. Because of size we suggest additional files for this.
• Why 3 models for Hospital and 2 for ER?
• ER: Please describe the scientific and the clinical significance of a) explained variance lower than 0.15 in men and b) difference in explained variance of 0.025 in men and of 0.076 in women.
• Hospital: see above for differences in explained variance between models.

Response:
• ‘Clusters related to aging’ are used in this modified version.
• We have loaded up an extra additional file. Additionally, we have named these variables together with the tables in this modified manuscript, so the readers with interest can get a quick browsing without opening the additional file.
• Compared with cluster morbidity clusters models (Model 2), no new variable was included in the cluster interactions models. That is to say, cluster interaction did not contribute to the models (one morbidity cluster was supposed not to be affected by another one).
• The variables selected in the models got a lot of our attention. A low $R^2$ (0.15) represents the models could still be explained by other factors. The minimal difference (0.025 in men and of 0.076 in women) but do a decreased $R^2$ informed us that morbidity clusters did not contribute impressively to the ER decision.
• In hospitalization models, $R^2$ increased when clusters and cluster interactions were involved in the models. The difference from ER is that the increased $R^2$ indicates an improvement yielded by the models’ parameters (morbidity clusters or cluster interactions).

Discussion
• Discuss the meaningfulness of reporting explained variances and differences in explained variance
• Do you interpret your models in such a way that all diseases not mentioned in tables 2 and 3 have no predictive value?: e.g.: stroke is not predictive for ER and Hospital? No disease is predictive for hospitalization in men, not even cardiac insufficiency or stroke? If true, discuss to which extent this might be due to clustering method (see above) or sample size or whatever? Propose reasons and/or alternative explanations. On the other hand, explain why GP contact becomes a predictor when people – according to table 1 – see a GP 1.5 times a year.

Response: We have added that in this new version (‘methodological issue’ and ‘multimorbidity patterns associated with ER visit and hospitalization’).

Low $R^2$ reminds us to concern different measure of behavior of health service use. The cases of first onset of these diseases (Stroke, heart disease, etc) are few at this age group. For very old people, these diseases were only their medical history rather than new diagnosis. Another hypothesis is about non-participants (with these diseases and in severe stages?). Number of GP visits was measured as a continuous variable. Every increased GP visit increases the corresponding odds of visiting ER or being hospitalized.

• Your explanations on p.8-10 are rather speculative. E.g.: “Very old adults might not be able to survive severe concurrent ailments (e.g., heart diseases) that are associated with a high risk for mortality”. Please explain more clearly this thesis as well as others.
Response: We have rewritten that in this new version. One reflection from daily clinical practice is that patients with a malignancy diagnosis usually have received complete clinical and laboratory examinations, so some comorbidity such as osteoarthritis and thyroid dysfunction would not be missed. Another hypothesis is based on the selection of survivals of concurrent ailments. Among cancer patients, some co-occurrences (e.g. severe heart disease) are more likely than others (e.g. osteoarthritis) to cause a high risk of mortality.

- P9: In terms of hospitalization, our results are consistent with other studies that multimorbid patients were more likely to be hospitalized [4, 28]. We do not see this in your data, please explain. And paper 28 did not investigate that question.

Response: A prevalence of multimorbidity was 68%. We added that in table 2. Furthermore, morbidity clusters were selected in the hospitalization models, so we interpreted that multi-morbidity had a higher likelihood. We agree with you and deleted Ref 28 in the current version. We previously only thought about the phenomenon of high multimorbidity shown in the work done by Freund et al (table 1).

- P 11. Going to see a GP 1.5 times a year shall be predictive for hospitalization? Please explain.

Response: Number of GP visits was a continuous variable. For every increased GP visit, the odds of being hospitalized increase by a factor of 1.5.

- P4: We did not understand the sentences needs to be concerned", "this trend is expected not only by chance", "It has been quite often that young elderly received more attention", P 8: „Gender stratification to some extent simplified gender’s comprehensive role in morbidity prevalence and related factors associated with health service use". Please paraphrase them.

Discretionary revisions

P4: We did not understand the sentences needs to be concerned", "this trend is expected not only by chance", "It has been quite often that young elderly received more attention", P 8: „Gender stratification to some extent simplified gender’s comprehensive role in morbidity prevalence and related factors associated with health service use". Please paraphrase them.
Response: In this modified manuscript, we have rewritten or simplified the sentences.

P.5: In your paper socioeconomic status is only represented by working status as measured by occupation. As you also measure education please do not call it “SES” but rather “working status”.

Response: We agree. ’Working status’ has been used in this current manuscript.

Minor Essential Revisions

Unappropriate wordings:
• P 7: “Cluster 5 where malignancy aggregating to osteoarthritis”, “linkage”, “primary care GP” (others do not exist), notable characteristics (did you skip the unnotables?), “adding greater specificity”
• P.8: “Model 3 was not structured”
• P 9: Comparatively, men had a somatic-mental cluster as only affective disorder was included in the analysis
• P.10: “However, to date it is not empirical method.”

Response: In this modified manuscript, we have rewritten them according to your advice. We want to remind readers that results from hierarchical clustering methods are not conclusive since very different results on the same data are recognized using different methods. No one method is recommended above all others [1] (p83). We have also realized that some other methods have also been used to measure multimorbidity [7, 8].

References


Response to
Dave Kerby

We greatly appreciate all your advice. We have made every effort to improve our manuscript. Below are the point-to-point responses to the queries. We hope that the responses have adequately addressed your concerns.

The response was done by all the authors together with statistician Ass.Prof. Mats Fredrikson (Linköping Academic Research Centre).

**Major Compulsory Revisions:**

1. The authors are not clear on how the presence of a chronic disease was measured. On page 5, the authors note that chronic disease was assessed in two ways: by self-report, and by medical records. But there is no mention of how these two sources were combined. The paper should state how chronic disease was defined when the two sources did not agree. For example, what if an older adult reports suffering from arthritis, but this diagnosis is not in the medical record? As another example, what if the medical record reports a diagnosis of dementia, but the older adult reports no memory of such a diagnosis? Also, the paper should state the nature of the self-report – whether it was a checklist, or a free response question.

   **Response:** In the current version, we rewrote this part (under the subtitle with ‘morbidity and health service use’).

   We shortly described how we did when self-reports did not match case reports: ‘A disease or condition was only registered if there was a clear documentation of the disease and its treatment, regardless of the patients’ self-reports.’

   One limitation is that arthritis sounded not a high prevalence in the study population, although it was included in both paper of John et al [1] and the Leiden 85-plus study [2]. We mentioned that in the discussion (under the subtitle with ‘limitations’).

   The information of predetermined list was also added. In practice, this was also part of postal questionnaire but enclosed several diseases as examples in each category.

2. On page 6 at the end of the first paragraph, the authors say, “we chose a prevalence >5% as the criteria for a common morbidity.” This statement describes how chronic disease was measured; as such it should not be in the data analysis section of the paper, but in the section describing how chronic disease was measured (see point 1 above).

   **Response:** We agree. We moved that to the method (under the subtitle with ‘morbidity and health service use’).

3. The authors are awkward in their description of the distance measure used. On page 6 at the top of the page: “A hierarchical cluster dendrogram was formed using Yule’s Q measurement.” In the context of cluster analysis, the usual term used is not “measurement” but “distance measure” or “similarity measure.” A better wording would be something like the following: “... was formed using Yule’s Q as the similarity measure between clusters, with a higher value indicating greater similarity.”
Response: With your good advice, we modified our descriptions in the current version.

4. The authors are imprecise in their description of the method of agglomeration. On page 6 on the third line, the method of agglomeration is referred to as “average linkage.” In fact, there are two types of average linkage: a) within groups, and b) between groups. Figure 2 states that the method used was average linkage between groups, so this should be made clear in the text. A simple statement of fact should do it, such as “the agglomeration method was average linkage between groups.”

Response: We modified our descriptions in the current version according to your suggestion.

5. On page 6, line 6: “A cluster feature can often be overrepresented when one or a [sic] just a few cases are investigated.” I am not sure what the authors are trying to say here. I am puzzled about the meaning of the word “cases”, because the cluster analysis reported in the paper is not of cases but of variables. I also am not clear on the meaning of “cluster feature.” The authors should drop this sentence or reword it in some way that clarifies the intended meaning.

Response: (please also see response No.2) we modified this part.

6. On page 7, in the second full paragraph: “A five-cluster structure was derived by a cut-off distance of 15 units. . .” The cutoff value based on visual inspection seems reasonable to me. However, there is a lack of clarity in the reporting, because the method section reported that Yule’s Q was the similarity measure. What the authors report here in the results section is the SPSS rescaled value, which has no intrinsic meaning. I strongly recommend that the authors report the Yule’s Q value that goes with the rescaled value of 15. This Yule’s Q value can be obtained by an inspection of what SPSS calls the agglomeration schedule. For example, consider Figure 2. Cluster 3 joins with Cluster 4 at a rescaled value of 15; the SPSS agglomeration schedule will state the Yule’s Q value at the point when these two clusters joined; this value will be in the column labeled “coefficient”, because Yule’s Q is the coefficient of similarity. I would also recommend that the odds ratio associated with this Yule’s Q value be reported. In this way, the cutoff value is stated in terms that will be understood by most readers – namely, the odds ratio.

Response: we rewrote and modified our descriptions in the current version. Your advice helped us improve our understanding of the method and interpret the results better.

In this manuscript, we loaded up new figures with a scale of coefficients calculated by Yule’s Q correlation matrix combined with average linkage between groups (shown in SPSS agglomeration schedule). Moreover, we added detailed legend (optional in BMC journal) and hopefully the readers can understand the figures without carefully reading the text. Odds ratio was reported with coefficient for the terminal nodes of each cluster. We wish that we have understood your advice.

Minor Essential Revisions:

1. The authors switch between the term “morbidity” and “chronic diseases”. The paper would be much easier to read if just one term was used, and my personal preference is “chronic disease.”

Response: thank you for your suggestion. We used ‘morbidity’ in the text (results and discussion) because we would like to be consistent with a few other researches as well as our earlier paper.

2. Table 1 reports inferential statistics, comparing men and women, but all that is reported
are the p values. I commend the authors for reporting exact p values for the non-significant results and not merely putting NS. In addition, I think it would be far more informative to have column three report the appropriate statistic (the t statistic, the chi-square, or the Mann-Whitney U) along with the p value. I would also prefer to see a fourth column with an effect size that goes with the statistics – the Pearson r for the chi-square, the rank-biserial correlation for the Mann-Whitney U test, and either Cohen’s d or the Pearson r for the t test.

Response: We rewrote the first part of result (under the subtitle with ‘completeness, representativeness, and sample characteristics’). We added the statistic method and exact p-value in the text as well as in Table 1 (we are not sure if we have understood correctly). The effect size was added as a fourth column in Table 1.

3. The paper reports five clusters for men and five clusters for women. I think that a paragraph or two at this point could compare the results with some previous studies. The clusters that emerged in this study of the very old seem similar to clusters that have appeared in previous research. A discussion of these similarities would be informative.

Response: We modified in this new manuscript by making comparison with three studies using the same method of analysis [1, 3, 4].

4. In Figure 1, one chronic disease is “osteoposis”, which needs correcting whether this is retained as a figure or converted to a table.

Response: thank you for your kind reminder. We made a new table according to your suggestion (see the next point) [1, 3, 4].

Discretionary Revisions:

1. Figure 1 does not work well for me, for two reasons. First, Table 1 does a good job of describing gender differences on a number of variables, so a second table (not a figure) to describe gender differences in chronic diseases seems sensible to me. Second, the focus of the study is on the topic of multimorbidity, not gender differences, so I do not see a need to highlight gender differences in a figure. Though figures are nice and often easy on the eyes, I would much prefer to see these data in a table and not in a figure.

Response: We agree that it was not necessary to emphasize ‘gender difference’, so we presented the morbidities as descriptive data. We made a new table (Table 2) instead. Moreover, we moved them to the result part since it was part of the characteristics for the population. We also modified the introduction.

References

