Author's response to reviews

Title: The psychometric properties of three self-report screening instruments for identifying frail older people in the community

Authors:

Silke F Metzelthin (s.metzelthin@zw.unimaas.nl)
Ramon Daniels (r.daniels@hszuyd.nl)
Erik van Rossum (e.van.rossum@hszuyd.nl)
Luc de Witte (l.de.witte@hszuyd.nl)
Wim JA van den Heuvel (heuvelwim@hotmail.com)
Gertrudis IJM Kempen (g.kempen@zw.unimaas.nl)

Version: 2 Date: 29 December 2009

Author's response to reviews: see over
Dear Dr. Kerr,

Thank you so much for your detailed comments on our manuscript. We do appreciate the efforts of the reviewers very much and we agree that relevant comments have been given to improve the manuscript.

In this letter we would like to give a point-by-point response to the concerns of the reviewers.

Reviewer #1
Discretionary Revisions (which are recommendations for improvement but which the author can choose to ignore)

1. Results in the abstract should be reported in the order which they appear in the methods section of the abstract.
   → Our response: We agree with the reviewer. We have made changes in the manuscript. In all sections (abstract, results and discussion) the results appear in the same order.

2. Demographics of the study sample could be compared with those from population-based studies in the study area.
   → Our response: We agree with the reviewer. We have added information on demographics of Dutch older people to the manuscript (page 12):
   The sample is representative for the Dutch population of older people. According to a report of the Netherlands Institute for Social Research (De Boer, 2006) slightly more people aged 75 years and older are female. Older people, especially women, are often less educated and have an average income of about €1500 (per month/per household).
3. Indication of what is generally considered to be acceptable Cohen’s Kappa or Cronbach’s alpha coefficients could be added.

   \( \Rightarrow \) Our response: To help readers interpreting the results we have added a rule of thumb regarding kappa values and alpha coefficients (page 13):

   According to Landis & Koch (Biometrics, 1977) the kappa values indicate good agreement between GFI and TFI and fair agreement between GFI and SPQ, and TFI and SPQ (<0.20 = poor, 0.20-0.40 = fair, 0.41-0.60 = moderate, 0.61-0.80 = good, 0.81-1.00 = very good agreement). Cronbach’s alpha coefficients for the GFI, the TFI and the SPQ were \( \alpha = 0.73 \), \( \alpha = 0.79 \) and \( \alpha = 0.26 \), respectively. The higher Cronbach’s alpha, the more reliable the test is. Alpha values above 0.70 indicate a satisfactory internal consistency for a scale (Nunnally, 1978).

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

1. More discussion of the definition of frailty and its relationship with disability and co-morbidity is needed. Some authors (Fried, J Gerontol A Biol Sci Med Sci 2004) consider these to be separate but overlapping concepts. The items in the frailty scales assessed and the decision to compare the frailty scales with a measure of activity restriction imply that the manuscript’s authors do not share this view. This should be clarified in the manuscript.

   \( \Rightarrow \) Our response: We agree with the reviewer that frailty and disability are separate but overlapping concepts. Therefore we have decided to add an instrument that measures disability more specifically (Groningen Activity Restriction Scale (GARS)). We have found a substantial correlation between frailty scores and GARS scores. That is an indication for construct validity. This has been clarified in the manuscript on page 11:

Frailty and disability are related concepts (Fried, J Gerontol A Biol Sci Med Sci 2004), as frail older people have an increased risk for disability and the other way around disability exacerbates frailty (Fried, J Gerontol A Biol Sci Med Sci, 2004; Levers, J Adv Nurs 2006; Markle-Reid, J Adv Nurs 2003). Substantial associations between frailty and disability were expected. Therefore construct validity was also assessed by examining associations between frailty and disability measured by means of the GARS (Spearman rank correlation). However correlations may not be too high. Otherwise frailty instruments and the GARS would measure the same concept.
Major compulsory revisions

1. Why it is important to do this study? What are the problems with the other approaches to operationalize frailty? There are two major approaches to handle frailty: frailty phenotype (or Fried's frailty) and the frailty index approach, see for example Rockwood & Mitnitski, J Ger Med Sci., 2007. It would be important to check any other measure of frailty against these two.

→ Our response: We have searched for frailty instruments that were developed from a multifactorial perspective on frailty and could be used in postal screening procedures. In the Netherlands the Groningen Frailty Indicator (Steverink et al., Gerontologist 2001) is an well-known and frequently used instrument, although little is known about its psychometric properties. Another instrument that is recently developed in the Netherlands is the Tilburg Frailty Indicator (Gobbens, J Am Med Dir Assoc, in press). For international comparison we added a well-known and feasible instrument for postal screening: the Sherbrooke Postal Questionnaire. Our goal was not to develop a new frailty instrument but to test the psychometric properties of available instruments that have been used in postal screening procedures. The frailty phenotype (mentioned in the manuscript, page 6) and the frailty index approach (added to the manuscript as an example for multifactorial instruments (page 6)) are indeed well known approaches to identify frail older people. However, as far as we know they are not used in postal screening, which was a selection criterion for our study.

We have added the following passage to the manuscript (page 3):

→ During the last decades various instruments, based on various definitions, have been developed to detect frailty. This has led to a diversity in prevalence estimates of frailty, ranging from 36% to 88% (Pel Little, J Nutr Health Aging, 2009). Little is known yet about the reliability and validity of these instruments and no gold standard exists. Therefore, more insight into the psychometric properties of frailty instruments is relevant for geriatric care and research in this area.

2. One problem with the suggested measures is that they identify from 40 to 60% people as frail while much smaller numbers are presented in the other papers. That should alert the authors. In addition, they found such big prevalence in community-dwelling people while much less prevalence was shown in more selected (not by their higher wellness) people. This makes me skeptical about the validity of the results presented in the manuscript.

→ Our response: We agree with the reviewer that prevalence estimates of 40-60% can be considered as high. We have added the following passage to the manuscript (page 15):

Prevalence estimates of 40-60% can be considered as high. One explanation may be that cut-off points proposed by the authors lack evidence from longitudinal studies (e.g., predictive power). Increasing proposed cut-off points would lead to lower
prevalence estimates of frailty. A recent literature review of Pel-Little and colleagues (Journal of Nutrition, Health & Aging, 2009) reports prevalence estimates ranging from 36% to 88% depending on the study sample and the instrument that is used. Our prevalence estimates are within this range. It is important to bear in mind that prevalence estimates strongly depend on the definition of frailty, instruments used and the target population. Longitudinal study designs are needed to obtain a better view on relevant cut-off points for frailty for various instruments and target populations related to risks of frailty in terms of, e.g., mortality, nursing home admission and (increased) disability.

3. I understand that the authors wanted to use a simple measure based on self reports. That is fine with the frailty index. It can be used (and has been used) in the self-reported setting. Perhaps the authors could consider this option if they chose to revise the manuscript.

→ Our response: The frailty index is indeed a simple measure as well, but – as stated above - we have not found reports that used the index in postal screening procedures. We therefore have not considered the frailty index as an option for this study. But it may be a point of interest for future research. We have added the following passage to the manuscript (page 17):

The frailty index (Searle, BMC Geriatr 2008; Rockwood, J Gerontol A Biol Sci Med Sci 2007) is a simple measure as well that is based on self-reports. However less is known about its feasibility for postal screening. To investigate the feasibility and validity of the frailty index as postal screening instrument may be a point of interest for future research.

4. Why does the proportion of frail people decrease with age (Table 1)? The same with the number of frail people with high income (see also Table 3). This contradicts to all what is known about frailty however it is defined.

→ Our response: Unfortunately, this may be a misunderstanding. Table 1 only gives a summary of the background characteristics of the study sample, it does not give information about associations between frailty and these characteristics. Therefore, table 1 renders no information on whether frailty increases or decreases with age. Table 3 reports differences in mean scores on the frailty instruments between groups with different background characteristics. Mean scores on frailty increase with lower income and higher age; both trends are in line with the frailty literature.

Minor Essential Revisions

1. BMC journals do not have space limitations.

→ Our response: We have chosen to display the items of the frailty instruments as additional files. However they may be well useful for readers who are searching for frailty instruments. If reviewers or the editorial team prefer to display the frailty instruments within the text, we are very happy to do so.
2. Tables should not be called ‘figures’.
   → Our response: We have denoted the tables ‘tables’ in the manuscript. However, we have uploaded them separately from the main text as ‘figures’. That’s why the term ‘figure’ is displayed in the left corner of the manuscript.

3. Reference to the paper just submitted is not appropriate (e.g. 18)
   → Our response: When we submitted our article to BMC Public Health in August 2009 the reference of the article of Gobbens and colleagues was unknown. In the meantime their article has been accepted by the Journal of the American Medical Directors Association. The complete reference is added to the reference list:
Reviewer #3

Required Compulsory Revisions

1. Provide more details about missing responses and the potential impact of missing data on estimates of internal consistency, agreement, and prevalence of frailty. For example, in the Discussion section, the authors note that the frailty prevalence estimate was 67% instead of 59% if non-respondents were considered at risk. What were the characteristics of non-respondents and the difference in characteristics between respondents and non-respondents. Furthermore, the single imputation strategy that was adopted will result in underestimation of the variability in the population parameters. The authors should adopt a multiple imputation strategy for their analyses.

Our response: As we have no information about non-responders, it is very difficult to assess the potential impact of missing data on estimates of internal consistency and agreement. We assume that a substantial part of the non-responders is either very frail (e.g. not able to fill in the questionnaire), or very healthy (e.g. they may think that the questionnaire is not ‘applicable for them’). Consequently, we expect that missing responses do not have a large impact on the prevalence estimates as frail and non-frail non-responders balance each other.

We stated in the discussion of the article that considering non-respondents as frail would lead to an increased prevalence of frail cases for the Sherbrooke Postal Questionnaire (SPQ) (59% → 67%). This increase depends on the SPQ protocol, in which non-responders have to be considered at risk (frail cases). As this ‘guideline’ is not adopted by the Groningen Frailty Indicator and the Tilburg Frailty Indicator, we decided to not use the non-responder rule for the SPQ.

Missing data is ubiquitous in survey data. There are different approaches and related strategies to deal with this problem. Each of the approaches and strategies has advantages and disadvantages. For the analysis of the validity of the chosen instruments we have chosen a single imputation strategy (mean case substitution) to handle missing values. Recognizing its limitations (e.g. underestimation of the standard errors or too small P-values, i.e., overestimation of the precision) we still believe that the chosen method is appropriate to handle missing values, as the single imputation method is a widely used approach and one of the most popular imputation strategies is mean substitution (Pearson, Mining Imperfect Data. Dealing with Contamination and Incomplete Records, 2005). Perhaps a multiple imputation strategy had lead to slightly other results, although it also may have led to biased outcomes (e.g. overestimation of the standard errors or too large P-values).

2. The use of item response theory to evaluate individual questions is an important step in the development of psychometrically sound instruments. Why was this method not adopted in the present study?

Our response: We have chosen to investigate the internal consistency based on a classical test theory perspective. Based on this perspective Cronbach’s alpha is a
useful coefficient for assessing internal consistency (Altman, BMJ 1997). Furthermore Cronbach’s alpha is well-known and easy to interpret, especially for readers with minor statistical knowledge. We therefore did not use the item-response theory. It may interesting if the item-response theory would reveal similar results. This may be a point of interest for future research.

3. The Kruskal-Wallis test appears to be used incorrectly. On page 10, the authors report that this test (along with the Mann-Whitney U test) was used to conduct pairwise comparisons of frailty scores. However, the Kruskal-Wallis test produces an omnibus test of equality of the distribution of scores for three or more groups. The authors should clarify the method used to conduct analyses of group differences.

→ Our response: We have used the Kruskal-Wallis test as non-parametric method to test whether the population means among groups are equal. We have not conducted pairwise comparisons. This was erroneously stated in the manuscript and is changed in the revised version (page 11):
Mann-Whitney U tests were used for pairwise comparison of frailty scores among males and females. Kruskal Wallis tests were used to test whether the population medians among groups with different participants’ characteristics (age, education and income) were equal.

4. The percentages that are reported beside each of the response categories should be removed because they repeat information that is already provided in Table 1. The authors should use a different method to report statistical significance. For each group comparison that was conducted, the test statistic value and p-value should be reported for each frailty instrument. For age, education, and income, the symbols * and ** could be erroneous if the Kruskal-Wallis test was used (see my note above) to conduct omnibus tests.

→ Our response: The percentages in table 3 have been removed. Test statistics and p-values have been added to table 3. The symbols * and ** have been removed from the table.

5. Report confidence intervals (i.e., 99% confidence intervals) for the Spearman rank correlations in Table 2.

→ Our response: Confidence intervals have been added to table 2.

Minor Essential Revisions
6. The Groningen Frailty Indicator, as described in Appendix 1, is written from a third-person. Were the items reworded to reflect a first-person perspective, consistent with the Tilburg Frailty Indicator and the Sherbrooke Postal Questionnaire?

→ Our response: The items of the Groningen Frailty Indicator are presented in the third-person in the publication of Schuurmans and colleagues (J Gerontol A Biol Sci
However, in the questionnaire we have used the questions in the first-person. The items of the Groningen Frailty Indicator have been adapted in the manuscript (appendix 1).

7. Table 3: all numbers should use a decimal point.
   → Our response: Two numbers in the table indeed had no decimal point. We have adjusted this in table 3

8. Information about the geographic location of residence (e.g., urban, rural) of study participants.
   → Our response: Two out of three general practitioner practices are settled in an urban area (Amersfoort and Roermond). One practice is settled in a rural area (Roggel). In total, 64% of respondents live in an urban area and 36% in a rural area. This information is now added to the manuscript (page 12):
   In total, 64% of respondents live in an urban area (Roermond, Amersfoort), while 36% live in a rural area (Roggel).

9. Table 3: the Sherbrooke Postal Questionnaire results in a higher score for males than for females, while the converse is true for the Groningen Frailty Indicator and the Tilburg Frailty Indicator. The authors should provide some explanation for this discrepant finding.
   → Our response: We agree with the reviewer that higher scores on the Sherbrooke Postal Questionnaire for males compared with females are inconsistent with the literature. We have no explanation for this finding. However other findings on the Sherbrooke Postal Questionnaire (higher score with higher age, lower educational level and lower income) are well in line with the literature (Avila-Funes et al., J Gerontol A Biol Sci Med Sci, 2008). A passage is added to the manuscript (page 16):
   Scores on the Sherbrooke Postal Questionnaire were higher for males compared with females. This finding is inconsistent with the literature (Avila-Funes et al., J Gerontol A Biol Sci Med Sci, 2008). However other findings on the Sherbrooke Postal Questionnaire (higher score with higher age, lower educational level and lower income) are well in line with the literature (Avila-Funes et al., J Gerontol A Biol Sci Med Sci, 2008).

10. Discussion section: the authors repeat some of the findings already reported in the Results section. This section should be limited to an interpretation of the study findings and a discussion of the strengths and limitations of the research.
   → Our response: We agree with the reviewer that the main part of the discussion should address the interpretation of the findings and strengths and limitations of the study. However we think that readers in general welcome a short summary with some main conclusions at the beginning of the discussion section. However, if the editor prefers to skip these, we are happy to do that.
Discretionary Revisions

11. While Cronbach’s alpha will produce the same result as the Kuder-Richardson Formula 20 for evaluating internal consistency, this may not be known to all readers. Some may mistakenly believe that Cronbach’s alpha is not an appropriate choice for dichotomous responses. An appropriate reference should be provided about their equivalence.

Our response: We have added a sentence to the manuscript (page 10):

A Cronbach’s alpha coefficient is calculated for (dichotomous) responses to explore whether items form one scale. In SPSS, the Cronbach’s alpha will produce the same result as the Kuder-Richardson Formula 20 (K-R-20), which is the alternative method to assess the internal consistency for dichotomous items (Kuder & Richardson, 1937).

Editor

Clarify if the study required ethical approval? If no approval was required, please let us know the circumstances under which ethical review was waived.

Our response: The study did not require ethical approval. In the Netherlands studies involving human subjects must undergo a medical ethics review, if they are subject to the Medical Research Involving Human Subjects Act (WMO). Completing a short questionnaire does not bring a study within the scope of the Act. Except if questions are highly personal, embarrassing or intimate, or if human subjects have to spend a considerable amount of time completing the questionnaire, which was not the case in our study.

From an ethical point of view we have paid attention that older people were well informed on the study in a patient information letter that accompanied the screening questionnaire. The letter was formulated according to the guidelines of good clinical practice.

We hope that our point-to-point reaction fulfils your expectation and that you will consider our manuscript for publication in BMC Public Health.

Yours sincerely,

Silke Metzethlin, MSc  Ramon Daniels,  MSc
Erik van Rossum, PhD  Luc de Witte, MD, PhD
Wim van den Heuvel, PhD  Ruud Kempen, PhD