Author’s response to reviews

Title: Development and validation of a measure of health literacy in the UK: the Newest Vital Sign

Authors:

Gillian Rowlands (rowlang2@lsbu.ac.uk)
Nina Khazaezadeh (Nina.Khazaezadeh@gstt.nhs.uk)
Eugene Oteng-Ntim (Eugene.Oteng-Ntim@gstt.nhs.uk)
Paul Seed (paul.seed@kcl.ac.uk)
Suzanne Barr (suzanne.barr@kcl.ac.uk)
Barry D Weiss (bdweiss@u.arizona.edu)

Version: 3 Date: 9 December 2012

Author’s response to reviews: see over
Dear editor

Thank you for providing these insightful referee comments. Please see our responses below. All the comments have been incorporated into the text except the response for point 5 from reviewer 1 (see below). We are happy to incorporate this comment into the text if you feel it would be helpful to do so.

Yours sincerely

Gill Rowlands on behalf of all the authors

Characteristics of people with low health literacy on coronary heart disease GP registers in South London: Response to reviewer comments

Reviewer (1): Marie-Louise Essink-Bot

Minor Essential Revisions
1. Abstract, Methods section: it is stated that 300 participants would be required for statistical significance. My question is: significance of which comparison? Is mentioning statistical significance appropriate in this context? Furthermore I suggest to mention in the Methods section that construct validity was assessed using ROC analysis, analyzing the association between NVS-UK scores and TOFHLA-scores.
Response: thank you for identifying this lack of clarity. The power calculation showed that n-250 would be required to identify a correlation similar to that in the validation of the NVS US original. This has been inserted. The text relating to the ROC analysis has been clarified. Also see response to point 3 below.
2. Methods section, paragraph on Sample and Recruitment: same comment. To me, the basis of the power analysis is unclear.
Response: see above and also response to point 3 below.
3. Methods section, paragraph on Data analysis: 
- I think that assessing correlations between NVS and TOFHLA is a useful analysis, but that it is not an assessment of criterion validity. Criterion validity means comparison with a gold standard; I do not think
that TOFHLA is a gold standard of health literacy. Correlational analysis may show the analyst the similarity of concepts being measured.

- I have problems with the word ‘predict’, as in predictive validity, and throughout the manuscript. There is no prediction of some future characteristics using characteristics measured now. This is a cross-sectional study; the researchers investigate associations between two characteristics measured at the same point in time. The association between NVS-scores and education, and the ROC analysis, are to be regarded as analyses of construct validity.

Response: thank you for these comments. The differentiation of different ‘types’ of validity is obviously unhelpful and has been removed. Instead we have described our principal validity analysis – comparison of the NVS-UK against an accepted measure of health literacy, the TOFHLA, and a secondary assessment of validity against educational level. The ROC analysis provides both an additional test of validity and enables identification of optimal cut-off point(s) to identify people with low vs. adequate health literacy. This has been clarified in the text.

4. Section Results and discussion, subsection Validation:
- please restrict numbers of decimal digits in SD to 1 (1.9 instead of 1.883).

Response: this has been changed

- Next paragraph: please indicate whether a ‘higher’ score on TOFHLA indicates better or worse HL.

Response: this has been addressed.

- Next paragraph: it is stated that the correlation against TOFHLA was acceptable ($r = 0.49$). What criterion did the authors use to judge acceptability of the size of this association? I think they may state that $r=0.49$ indicates about 70% of shared common variance, which seems to me a large proportion, indicating similarity of measured concepts.

Response: We have revised the wording of the power calculation (as below) to clarify why we regard 0.49 as “acceptable” in this context; and amplified the reporting of the results by adding confidence intervals and a significance test against 0.3.

Rewording:
(Methods) “An unacceptable correlation was considered to be 0.3 (i.e. accounting for 9% of variance, and (based on previous data) a plausible correlation for purposes of power calculation was defined as 0.5 (or more). A correlation is therefore regarded as acceptable if it is shown to be significantly higher than 0.3.”

(Results) “The correlation against the reference standard TOFHLA was 0.49 on 332 observations (95% CI: 0.40 to 0.57), meaning that 24% (95% CI 16 to 32) of variance is accounted for; which can be deemed acceptable as it is significantly higher (P<0.001) than the unacceptable value of 0.30 set in the power calculation.”

- two paragraphs below: I suggest to add some explanation to the figures about sensitivity and specificity. E.g. 40% specificity in this context means (if I understand the matter correctly) that 40% of the subjects with adequate health literacy as defined by TOFHLA also have a good score on the UK-NVS; i.e., 60% of subjects with adequate HL according to the TOFHLA obtain an inadequate score on the NVS. This may indicate 60% false-positives; or it suggests that the NVS is more difficult, i.e., more strict than the TOFHLA.
Response: thank you for this comment. We have expanded both the results and discussion sections in relation to this. Neither the NVS-UK nor the TOFHLA measure all aspects of health literacy or even identify a specific literacy or numeracy level. It is therefore not possible to say which is the ‘better’ or ‘stricter’ test. What can be said is that the NVS-UK is quicker and easier to apply than the TOFHLA and also that it’s lesser skewness means that it measures health literacy skills across a wider range than the TOFHLA.

5. Discussion
- I invite the authors to add their comments on the issue of comparability of UK and US NVS. Can we assume that UK NVS scores are measured on the same scale as US NVS scores; i.e. if a UK population has an average UK NVS score of 3 and a US population has an average US NVS score of 3, can we conclude that the HL levels of
these populations are comparable? If this can not yet be assumed, what additional research is necessary to establish international comparability of scores?

Response: This is an interesting question. Given that the NVS UK was directly derived from the US original, and we attempted to preserve the level of difficulty of the NVS questions, it is likely that people with similar scores in the US & UK have similar levels of literacy and numeracy as applied to health; however it would not be possible to be sure of this without direct comparison i.e. giving both to the same subjects under similar circumstances. Such a study would have to involve subjects who were equally familiar with US English and food labels, and UK English and food labels; and moreover had a generally low level of literacy - a very difficult group to recruit to. The other aspect to consider is that absolute levels of health literacy are only half of the issue; the other being the complexity of health information and health systems, which are likely to vary considerably between countries.

The authors feel that whilst this is an intellectually interesting argument, it is unlikely to be of practical importance in UK and US research and practice. We have not, therefore, inserted a paragraph on this into the discussion but are happy to do so along the lines outlined above.

- paragraph above ‘Conclusion’ section: The sentence starting with ‘Finally, although …. ’ gets lost – some words wrong or a typo somewhere? Please correct.
Response: thank you for identifying this; the typographical error has been corrected and the wording clarified.

Reviewer (2): Joanne Protheroe
Minor Essential
My main amendments would be to the methods chapter.
1. There is a lack of detail in some of the sections pertaining to recruitment. It is fairly clear how the panel of experts were recruited for
the Delphi technique, presumably on-line (although this is not stated).
Response: this clarification has been added.

2. It is less clear how participants were recruited for the cognitive testing or validation studies. For the former, the authors state that the participants were recruited purposively from Lambeth residents – but not how, face to face on the street, postal? Nor do they describe how they purposively sampled according to their criteria.
Response: this has now been clarified.

3. In addition it is not clear to me, not having any knowledge of ‘market research’, what grades D and E represent, a little more description would benefit the reader.
This has now been clarified in an additional table, and a reference added.

4. For the validation study there is no description of the recruitment process beyond post codes and age bands – where did they get this data and then how were participants invited to take part? In addition I would like to see more detail of how the authors arrived at their sample size calculation.
Response: more details about the recruitment strategy have been added. The sample size calculation has been clarified (see response to reviewer 1).

5. In the results section it would be interesting to see who made up the recruited expert panel that were approached to take part from clinical practice, public health, dietetics, research, adult education and food and drink industry.
Response: this has now been added both as text and as an additional table (table 2). All other tables have been renumbered. We have identified a typographical error in the original text: 28 experts participated in the Delphi study. This has been amended in the text.

6. The results are well presented, although I found Table 1 and Table 2 slightly difficult to interpret and these would benefit from clearer labelling, so that the reader does not need to keep referring back to the text. In Table 1 I presume that the figures under the column headed Lambeth mean the % of the Lambeth population that is made up of the row characteristic. This is not clear. In Table 2 the insertion of an extra row for ‘Sample’ makes things a little clearer, as do the % signs in the
column headings, but still a little confusing.
Response: these tables are now 3 and 4. Table 1 (now 3) has been amended as suggested. Additional explanatory notes have been added to make the tables self-explanatory and avoid the need to refer back to the main text.

7. The grammar and standard of writing is excellent with very few typographical errors:
- Missing word page 6 ‘data were also collected socio-demographic’
  - Response: this has been corrected
- Incorrect word page 7 ‘participants were from groups like to have low health literacy’
  Response: this has been corrected