Reviewer’s report

Title: Development and testing of the KERNset: an instrument to assess the quality of telephone triage in out-of-hours primary care services

Version: 0 Date: 10 Aug 2016

Reviewer: Molly Jeffery

Reviewer's report:

Thank you for the opportunity to review this paper describing the creation and validation of an instrument to measure the quality of telephone triage calls. The authors requested input from quality experts at each GP cooperative organization in the Netherlands to determine the concepts that should be included in measuring quality of triage calls, assessed multiple existing triage quality questionnaires for questions measuring the key concepts, and convened an expert panel to assess the suitability of each of those questions and choose 24 to be included in the instrument. They selected a sample of 120 (reduced to 114 usable) triage calls, trained 8 raters, and assessed inter- and intra-rater reliability, item distributions and correlations with each other, and internal consistency.

The authors state that there are several triage quality instruments currently in use in the Netherlands and that it would be useful to have a single minimum set of questions that are the same across all groups measuring triage quality. The procedures used to develop and validate the instrument seem to have been appropriately designed to achieve the stated aims.

In general, the writing is clear and concise. A few key phrases seem to have lost some clarity in translation to English. I have noted those items in the attached marked-up copy of the MS.

The remainder of this review is divided into major and minor comments.

Major comments:

* The paper would be strengthened by adding a comment on the definition of "quality" of triage calls. The meaning is suggested by the choice of questions on the instrument, but an explicit discussion would be a good addition to the paper. It is apparent that there are two major domains: medical and communication. A brief discussion of what would constitute high quality in each of these domains would be helpful. This is especially important because the background section presents two known associations between characteristics of triage calls and quality of those calls that are later used to validate the instrument, but the information on what constitutes quality isn't presented until the results section.
The use of these known associations between characteristics of triage calls and quality of those calls as a validation exercise is not properly justified. The authors state that calls that are longer and calls for more urgent problems have been shown to have lower quality in prior studies using other instruments to assess quality. The fact that the new instrument replicates this finding does not seem to me to be very interesting, since the new instrument was created using items (presumably) taken from those other instruments. It is unclear (and not explained) why more urgent calls tend to be of lower quality; I can imagine a situation where more urgent calls are measured as having lower quality because the triagist skips some parts of the triage script when it becomes clear that the patient needs immediate care and those skipped questions are unlikely to change that rating. In this case, a measure that doesn't take that into account would be inaccurately assessing quality. The fact that the new instrument replicates this inaccurate assessment doesn't improve its validity compared to the other instruments.

One item that was frequently marked "N/A" was dropped from the instrument after this analysis. I'm not convinced by the reasoning for dropping it. Item 12 seems to be one of the most important questions in its impact on patient safety. The fact that it measures a rare event is not an adequate reason for dropping it.

Minor comments:

There is inadequate sample to comment on the ability of the instrument to discriminate among triagists. There are only three calls per triagist, and they are not a representative sample.

In some of the analyses, items rated "N/A" were treated as missing. An argument could be made that items that are "justly absent" (the criterion for marking an item as "N/A") have been "performed optimally, with no room for improvement" (the criterion for marking an item 4)

Two questions were dropped for being strongly correlated with another question. However, the sampling design doesn't yield a representative sample of calls (more urgent calls are over-sampled and the calls for a single triagist had to have different chief complaints). I'm not sure whether you can assess item correlation in these circumstances.

Purists would disagree with the choice to treat an ordinal scale as continuous (i.e., in reporting means and in the regression analysis). It would be worth commenting on this choice.
Are the methods appropriate and well described?
If not, please specify what is required in your comments to the authors.

Yes

Does the work include the necessary controls?
If not, please specify which controls are required in your comments to the authors.

Unable to assess

Are the conclusions drawn adequately supported by the data shown?
If not, please explain in your comments to the authors.

Yes

Are you able to assess any statistics in the manuscript or would you recommend an additional statistical review?
If an additional statistical review is recommended, please specify what aspects require further assessment in your comments to the editors.

I am able to assess the statistics

Quality of written English
Please indicate the quality of language in the manuscript:

Needs some language corrections before being published

Declaration of competing interests
Please complete a declaration of competing interests, considering the following questions:

1. Have you in the past five years received reimbursements, fees, funding, or salary from an organisation that may in any way gain or lose financially from the publication of this manuscript, either now or in the future?

2. Do you hold any stocks or shares in an organisation that may in any way gain or lose financially from the publication of this manuscript, either now or in the future?

3. Do you hold or are you currently applying for any patents relating to the content of the manuscript?

4. Have you received reimbursements, fees, funding, or salary from an organization that holds or has applied for patents relating to the content of the manuscript?

5. Do you have any other financial competing interests?

6. Do you have any non-financial competing interests in relation to this paper?
If you can answer no to all of the above, write 'I declare that I have no competing interests' below. If your reply is yes to any, please give details below.

I declare that I have no competing interests.

I agree to the open peer review policy of the journal. I understand that my name will be included on my report to the authors and, if the manuscript is accepted for publication, my named report including any attachments I upload will be posted on the website along with the authors' responses. I agree for my report to be made available under an Open Access Creative Commons CC-BY license (http://creativecommons.org/licenses/by/4.0/). I understand that any comments which I do not wish to be included in my named report can be included as confidential comments to the editors, which will not be published.

I agree to the open peer review policy of the journal