Reviewer's report

Title: The Queensland High Risk Foot Form (QHRFF) - Is it a reliable and valid clinical research tool for foot disease?

Version: 1 Date: 19 June 2013

Reviewer: Anita Raspovic

Reviewer's report:

Major Compulsory Revisions

Abstract, results: The 1st sentence suggests the reliability (as well as validity) is moderate to perfect however the lower reliability statistic is 0.41. Arguably terms to describe the strength of such relationships can be somewhat arbitrary although of general use to assist interpretation by an audience unfamiliar with these statistics. It could be argued that a description of moderate is too generous for a correlation of 0.41, particularly when the context of the relationship is to justify a clinical test. It is suggested that the authors consider the suitability of these descriptors which would be well placed into context by the addition of an absolute reliability statistic (i.e. one that quantifies the average difference between scores in the units of measurements of respective items) such as 95% limits of agreement.

Methods, phase two: What part of the patient assessments were blinded? Please explain a little more. In addition, at this stage more reference to the QHRFF is being made however the reader hasn’t been told much detail about the tool itself. It would be useful to understand more about the QHRFF at this stage along the lines of the last paragraph in the phase one results section (pg 14).

Methods, settings & participants: Paragraph 1 – Patient participants were recruited from high-risk foot services and had current or previous ulceration however the tool is described as being developed to detect high-risk foot factors and foot diseases in general populations. How does this sample represent a general clinical population as the first sentence of the discussion states? In addition, if clinician participants knew the patient participants were drawn from high-risk foot services could this have biased their measurements inadvertently? Were any of the participants known to the clinicians through past clinical consultations?

Methods, procedures: This section is difficult to follow in areas particularly as several steps are discussed together. It would be beneficial for the readership if this section was re-worked to describe exactly what was undertaken, step-by-step, for each discrete part of the evaluation of the QHRFF. This might even be broken into sub-sections with headings such as: criterion validation, inter-tester reliability and intra-tester reliability. Exactly what was undertaken for each step should be presented in enough clarity and detail that another
researcher could replicate the study based on the methods described in the paper.

Statistical analysis: The last paragraph presents one system of categorisation for Kappa and ICC values (which is used in the results and discussion substantially). Given the context of the data is vital to the suitability of such categories, it is suggest that the selected system is too generous in its ratings. For example, I would not consider an ICC of 0.41 as constituting moderate agreement when the purpose of the measure is of a clinical assessment tool. Readers with less understanding of these statistics may not understand the strength of the relationships found and may rely substantially on these descriptors. In addition, an absolute reliability statistic (i.e. one that quantifies the average difference between scores in the units of measurements of respective items) would be highly valuable to provide a true sense of the magnitude of differences found. Given this system is used substantially to derive the study findings, the researchers are asked to comment on this issue.

Minor Essential and Discretionary Revisions

Abstract: The wording in the abstract would benefit from some fine tuning. For example; foot disease does consume considerable health care resources, the word ‘appears to’ makes it sound like you are not sure about the statement. This is the same for sentence 3 where ‘appear’ is used. I found the use of the terms foot disease, foot disease complications and high-risk foot factors somewhat confusing, perhaps because they are being used in a specific way in this paper that is new to me. The terms ‘foot disease’ and ‘high-risk foot complications’ (throughout the paper) might make more intuitive sense to a broader readership however this is at the discretion of the JFAR editors and authors of the paper.

Abstract, Methods: 4th sentence – This sentence reads as if the clinicians were also attending the high-risk foot service, re-wording is suggested. Also, it is not clear why three cohorts of patients were recruited and how many patient participants there were per cohort.

Abstract, conclusions: 2nd sentence - Will recommendations for alterations to the tool be made for subsequent versions? Also, I suggest that use of the word ‘appears’ in the last sentence be re-considered.

Background: Paragraph 1, 1st sentence - Would you say that foot disease is the cause of death (i.e. leads to) or are they are strongly associated due to co-morbidities?

Background: Paragraph 2, last sentence - the word ‘seems’ sounds a bit unconvincing. In addition, this sentence talks about analysing and measuring high-risk foot services and patient outcomes although so far the QHRFF has been presented primarily as a tool to assess disease. This sentence might read better if it directly reflected the fundamental purpose of the tool as presented in this paper.

Background: Paragraph 3, 1st sentence –I found this confusing to follow given
the terminology used. Perhaps break the sentence into a couple of sentences and simplify.

Methods: Paragraph 1, 3rd sentence – was informed consent obtained from patient and clinician participants? I see this is stated later but would be worth clarifying here also.

Methods, phase one: Sentences in the final paragraph are a bit repetitive within the paragraph and with other paragraphs in this section.

Methods, settings & participants: Paragraph 1 - Why were 3 patient cohorts used instead of all aspects of tool evaluation being performed on a larger sample size? What was the sample size per cohort?

Methods, procedures: Paragraph 4 discusses the criterion measure of diagnosis made by an expert clinician. How robust is this measure as the gold standard diagnosis compared to respective qualitative testing? What is meant by testing the independence of the criterion measure?

Methods, procedures: Paragraph 5, which clinician data were compared for inter-rater reliability? General clinicians with each other?

Methods, procedures: Last paragraph - Between one and four weeks apart is a relatively inconsistent time period for re-measures given issues such as clinician memory of test results. Can you more specifically state which time period it was?

Statistical analysis: Paragraph 2 - states that all QHRFF items were tested for validity and reliability however on pg 9 it states that only QHRFF items relating to the foot disease construct were tested. Please clarify.

Discussion: 1st paragraph - At what point (by what measure) is it decided that items require revision and will these be addressed in the future or not? Please specify.

Table 1: Is this table essential given the large amount of data in subsequent tables? If so can it be condensed for presentation?

References: Are all of these references necessary / referred to in the paper? This list would benefit from being condensed as it is long and unwieldy. Please check JFAR protocols.

Level of interest: An article of importance in its field

Quality of written English: Not suitable for publication unless extensively edited

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

Declaration of competing interests:
I declare that I have no competing interests