Author’s response to reviews

Title: Exploring the reliability and acceptability of cognitive tests for Indigenous Australians: A Pilot Study

Authors:

Kylie Dingwall (Kylie.dingwall@menzies.edu.au)
Allison Gray (agray@uon.edu.au)
Annette McCarthy (annette.mccarthy@menzies.edu.au)
Jennifer Delima (Jennifer.delima@nt.gov.au)
Stephen Bowden (sbowden@unimelb.edu.au)

Version: 1 Date: 07 Apr 2017

Author’s response to reviews:

Dear Ms Clark:

Thank you for the opportunity to resubmit to your journal. Please find attached revised manuscript and see below for responses to reviewer comments.

Reviewer 1 Comments:

The paper has no falsifiable hypotheses (the abstract's Conclusion states that all 4 tests tested "showed the highest reliability"), nor does it seek to perform analyses or evaluations of validity using available tests

• The aim of the study was to describe the test-retest reliability and acceptability for four cognitive tests. Testable hypotheses are implicit in the reported analyses, for example, that the reliability correlations are significantly different from zero. Explicit hypotheses were deemed unnecessary to meet the stated aim. This article did not aim to assess validity and makes no claims as to the validity of the tests. It does however state that this should be a topic for further research on page 29.
It is unclear how the sample was selected; it sounds as if this is a convenience sample of hospital patients who happen to have taken certain tests and later provided consent.

• the study utilizes a convenience sample who were selected prospectively based on stated inclusion and exclusion criteria. It is not a retrospective study. Additional detail has been added to address this concern within the method section (see p.6 of revision).

The tests have nothing to do with WKS, yet the convenience sample includes patients with this condition. It is unclear how the particular decrements in WKS are being picked up by the test battery.

• This is a pilot study seeking to examine the reliability of four cognitive tests prior to their use in an RCT with WKS patients. This study excludes participants with any currently identified neurological conditions. The sample was selected to best reflect the conditions of those who would be recruited to the main study in order to describe how the tests perform in terms of reliability with in this sample. An introductory sentence was added to the methods about test selection and specific WKS clinical features measured (page 7). There is more detail provided about the domains measured under the individual test methodology.

Readers have no way of knowing whether the sample represents most Aboriginals.

• It was not the intent of this study to recruit participants representative of ‘most Aboriginals.’ As stated the intent was to recruit a selection of Aboriginal people who were representative of those likely to be recruited to the main RCT to follow. As stated in the methods, the sample included people who spoke a range of different languages, were of a range of ages over 18 and who were not pregnant (as this was a condition of inclusion for the main trial). Aboriginal Australians are highly diverse peoples, with many different cultures, customs and languages. This is alluded to within the Participants paragraph of the Methods. The participants are generally representative of the geographic region of Central Australia where the study took place. To address the concerns of the reviewer and for an audience that may not be familiar with Indigenous Australians an additional sentence was added to explain the context (page 6).

Introduction. Authors claim several times throughout the Introduction and manuscript that mainstream cognitive tests are based on Western concepts. This is false and makes it sound like authors are oddly scapegoating an entire region of the world; mental status testing has deeper roots in ancient China than anywhere in the Western hemisphere or Europe.
• We are not seeking to comment on the history of psychological testing, but are attempting to point out that the tests currently used in western, English speaking countries such as Australia are commonly based on language and concepts relevant to Western culture. Such ‘Western’ concepts are commonly not compatible with non-Western views of the world, and therefore tests need to be examined for their reliability when used with non-western cultures such as Aboriginal Australians. The sentence has been amended to clarify this point (page 5).

The second paragraph is a non-sequitur about retest effects. Authors never evaluate retest effects in their sample; all they report is test-retest reliability.

• Retest/learning effects are evaluated and presented in the results section See table 3 (page 41) and discussed in the Discussion (page 26).

Authors state that Indigenous people are the only ones who are able to determine relevance and acceptability of cognitive tests.

• There are few normative samples in cognitive tests based on Aboriginal people. Feelings of powerlessness as a result of dispossession and colonisation are still felt by Aboriginal people in Australia. It is imperative that Aboriginal people be consulted and involved in decision making regarding their healthcare as a means of self-determination. Sentence on page 5 has been amended to state Aboriginal people are “best placed to determine relevance and acceptability of cognitive tests.”

Methods. Why were pregnant participants excluded?

• These patients were excluded as there would not be any approached for the RCT. It was intended that the pilot characteristics would reflect RCT conditions.

Why did authors select these particular tests?

• Tests were selected based on the cognitive abilities measured in relation to WKS. An additional sentence has been added to page 7 of the methods explaining this. The individual test descriptions also outline specific cognitive abilities assessed.
How many raters were used to score the subjective elements of tests like story recall, and how were discrepancies handled?

- There were two raters and an average of results used. This was specified in the statistical analysis section. This has now also been clarified in the Story Recall method on page 10. An additional sentence has also been added on page 9, to the RUDAS method to discuss the subjective element of the test.

Authors must make clear how the qualitative research in their convenience sample differs from a typical English-speaking sample.

- We have utilised a phenomenological approach which seeks to describe the essence of the participant experience without inserting researchers’ theories or previous research. The data reported highlights that there are similarities and subtle differences between the study population and an English-speaking sample.

  o Issue with “Older people would struggle using the computer…”: the quote provided highlights Aboriginal peoples’ perception on differences in education. According to the Australian Bureau of Statistics, in 2006, approximately 45% of non-Indigenous people reported finishing year 12 compared to 19% of Indigenous students [1].

  o Issue with “My Husband a whitefella and I understand English”: this quote demonstrates the importance of language. The participant understood English and perceived the test to be easier.

There is nothing here but speculative text based on authors' opinions.

- With respect, two of the three reviewers could see the relevance of this study as there are minimal test-evaluation or psychometric studies in this population, as noted in our Introduction.

Reviewer 2 Comments:

It would be good to know what treatments the subjects were undergoing during this period and to discuss whether this could have influenced test-retest reliability, e.g. patients may have improved due to medication etc.

- A new Figure with medications has been incorporated into the text. See separate attachment as per submission guidelines.
It would be better to use ICCs instead of Pearson correlations to analyse test-retest reliability

- We have redone the reliability calculations using ICC consistency and agreement as suggested. See table 4.

I would like you to discuss reasons for attrition in more detail and why it was different for the various tests. 43% were lost for SMRT. Was that because subjects refused? At some point you should mention reasons for drop out. Maybe provide a flow chart.

- Difference in attrition rate has been incorporated into the text on page 6.

Please mention inter-rater reliability for SMRT in the abstract

- This has now been added to the abstract as suggested.

Though effect sizes are small and it probably does not make a difference, I suggest using non-parametric tests for comparing baselines and retest scores give the small number of subjects.

- As now stated in the manuscript on page 26, “to ensure that the pattern of statistical findings was not affected by distributional violations in small samples, we re-ran all the parametric tests of mean difference with a non-parametric test (Wilcoxon signed rank test), and found the same pattern of results. Hence only the parametric analyses are reported.”

Some quotations of patients are repeatedly given. Though they may have been linked to several topics, these repetitions may be avoided.

- Duplicated quotes have been removed from the manuscript.

Reviewer 3 Comments:

Main text and Table 1 did not include any information about patient diagnosis and condition. It is critical information for readers to examine significance and limitations of this study. Particularly, if their diseases include acute phase neurological disease, short-term reliability would be lower than expected. Please describe inclusion and exclusion criteria and clarify those points.
• Participant diagnoses and conditions are now described in a new Table 2. To clarify, those with neurological conditions were excluded from analysis. Inclusion and exclusion criteria are described on page 6.

Generally, test-retest intervals for reliability assessment range from a week to a month. To assess test-retest reliability, 1-5 days interval seems too short. Please explain the rationale for this short interval.

• The retest interval has been addressed in the text of the article on page 11. As this was a pilot study for an RCT taking place in the hospital environment, it aimed to replicate the conditions of the RCT. Also, a patient taking their own leave is an issue to be considered in the context of hospital stay. There is often little opportunity for follow up as many patients reside in very remote communities of Australia.

ICC was used to check only inter-rater reliability for SMRT, but ICC is also useful to check intra-individual test-retest reliability for all four cognitive tests.

• This has been incorporated into Table 4.

In Acceptability Interview Analysis, the phenomenological approach was employed for analysis of interviewed data.

Rationale of the approach should be cited in the section because phenomenological approaches could differ according to its theoretical background.

• The method of phenomenological approach has been added to page 12 within the Procedure section.

First language of a large part of the patients was not English. As the authors describe shortly in Discussion section, please describe the reasons to use the English version in Introduction or Methods section.

• This has been addressed under Materials/Apparatus section on page 7. Provision of healthcare in Central Australia is largely conducted in English. The tests selected were not yet translated into Aboriginal languages of the region. As a result of this study, the most appropriate cognitive tests were translated into several key languages for use in the RCT.
Did the participants understand test instructions properly? If the authors checked it in the interview, please show the information

- This is briefly mentioned in the limitations of the study on page 30.

In page 17, results showed that patients were interrupted or distracted for some reasons. Did those interruption or distraction affect the test performances?

- It was stated in the qualitative results section that it was observed that distractions may have impacted on performance (see page 19). This issue was also explained in the discussion.

In page 24, the sentence 'Interview data suggested that the tests demonstrated face validity as they were viewed as a good way to "test the brain"' include patient's verbatim, but I cannot find such description in Results section.

- The quote has been replaced with other participant quotes that are listed in the Results on page 27.

Please discuss about wide-range of confidence intervals in correlation coefficients between baseline and retest, which may be due to limited number of patients who participated both baseline and follow-up assessment.

- As now stated on page 26 the wide confidence intervals are a function of small sample size.