Author’s response to reviews

Title: Non-Invasive assessment of vibration perception and protective sensation in people with Diabetes Mellitus: Inter- and Intra-rater reliability

Authors:

Sean Michael Lanting (sean.lanting@newcastle.edu.au)

Martin Jeremy Spink (martin.spink@newcastle.edu.au)

Peta Ellen Tehan (peta.tehan@newcastle.edu.au)

Stephanie Vickers (stephanie.vickers@uon.edu.au)

Sarah Louise Casey (sarah.casey@newcastle.edu.au)

Vivienne Helaine Chuter (vivienne.chuter@newcastle.edu.au)

Version: 1 Date: 06 Jan 2020

Author’s response to reviews:

REVIEWER 1

Many thanks for submitting your paper describing your worthy research project. This is a nice little study providing some valuable information pertaining to current practice.

Author: Thank you for taking the time to read and review our research article.

Comment 1: I have little to recommend in the way of amendments, although I do think it would be helpful for some readers to clarify / define what is meant by inter- and intra-reliability as not all readers will be able to readily distinguish between the terms.

Author: Thank you for this comment, we agree that not all readers will be able to distinguish between inter- and intra-rater reliability without clarification. As such we have added the following sentence to the Background section, “Reliability refers to the level of consistency of measurement results between different clinicians (inter-rater) and the same clinican on multiple occasions (intra-rater)”.

Comment 2: You have highlighted the low numbers in your limitations, but have you any plans to try and repeat with more participants?

Author: Thank you for your comment. The research was completed as part of Honours projects on a relatively short time-frame and thus the smaller numbers. At present, the sample size will remain a limitation of the study however there is scope for future research in this area.

Comment 3: I have noticed that some of your references are quite old. Only 7 of your 38 references are less than 10 years old. Perhaps this is worthy of a comment within your discussion - perhaps confirming the paucity of data around this topic?
Author: Thank you for your comment and you are correct regarding the paucity of recent evidence in this field and hence the need for research. I can confirm that the search for literature was quite exhaustive; primarily as a literature review was also undertaken at the same time.

REVIEWER 2
Thank you for the opportunity to review your article. This was well-written and useful body of work. I feel this is of a standard suitable for publication. I have some minor comments below.

General comments:
Possibly the biggest weakness in the study is that participants are considered as having DPN or not based on a written diagnosis in the medical records. The validity of this is questionable. If this diagnosis could have been confirmed via nerve conduction studies or another valid technique this would have added more depth to the study and enabled sensitivities and specificities to be calculated (and provided greater understanding the role practitioner experience may have had). However, this of course does not detract from the ability of the study to determine inter and intra-rater reliability.
In an ideal world, greater numbers with a broader representation of type of diabetes would have also been good.

Author: Thank you for taking the time to review our research article. We agree with your suggestions regarding the usefulness of a larger study design and comparisons with nerve conduction studies. This project unfortunately was undertaken to investigate reliability only and thus in this instance, was not undertaken.

Comment 1: Considering you comment a number of times that multiple tests should be used (rather than relying on one test to diagnose DPN), is it worth looking at this? I.e. make a diagnosis (for those which had multiple tests) and compare between raters. Does this increase the reliability?

Author: Thank you for your suggestion. This paper is an investigation into the reliability of some instrumentation and methods used in a population in which the tests are commonly used. The statements that these screening tests ideally should not be used in isolation stands true, based on the study findings coupled with that of existing literature. As you have rightly pointed out the reliance on diagnosis of DPN via medical records does not reflect a gold standard reference test that could be used to investigate the diagnostic accuracy of index tests such as monofilament or neurothesiometer which measure different aspects of neurological function. Therefore while the test with the highest reliability individually would have been the tests of choice to use in combination for maximum reliability this does not reflect the diagnostic accuracy of the test individually or in combination. This has been clarified in the discussion of the paper and highlighted as a requirement for future research. “Of note, our results relate specifically to the reliability of the tests used, i.e. that the results can be replicated, not that they reflect a correct diagnosis of DPN. While use of tests with high reliability is essential for effective clinical management, so too is the need for the tests to be able to diagnose the target condition. It has been stated that two-test combinations have >87% sensitivity in detecting DPN (36), though further work to determine the combination test with highest reliability that is most diagnostically accurate for identifying presence of DPN is required”.

Background
Comment 2: Page 3, Line 37 - It would be good to update this reference (and associated detail) to reflect the most recent guidelines released by IWGDF.
Author: Thank you for your recommendation. We have updated the document to refer to the 2019 guidelines in place of the 2015 guidelines.

Comment 3: Page 3, Line 40 - Specify 10g monofilament.

Author: Thank you for your recommendation, we have now added ‘10g’, to the sentence.

Methods

Comment 4: Page 4, Line 36 onwards - More information about how the flyers were distributed, or the exact setting of the recruiting clinic could be provided (i.e. in a hospital, university-based, GP centre)

Author: Thank you for your recommendation. The location was specified in the paragraph above, however we have now added more detail to the sentence, specifically, “with flyers posted up in university clinic consultation rooms and the waiting room, directing potential recruits to register their interest”.

Comment 5: Page 5, Line 37 - Given the varied experience of the assessors it would be good to clarify the technique of monofilament application (i.e. applied until the monofilament bent, held for X seconds and removed) or at least confirm that consistency between examiners was ensured.

Author: Thank you for your suggestion. We have added the following sentence to satisfy your request, “Monofilaments were applied perpendicular to the skin until buckling and held in place for 1-2 seconds”.

Results

Comment 6: Page 8, Line 4 - ‘substantial’

Author: Thank you, typo now fixed.

Comment 7: Page 8, Line 14 - 20 - Given that the results for intra-rater reliability is presented in the format of the ranges for all raters, it would be good to specify that this data (i.e. К=0.61 etc.) relates specifically to the reliability between all three raters (rather than the ranges calculated when comparing individuals).

Author: Thank you for your suggestion. In a reliability study, we feel it is important not to sound too repetitive to allow for easier readability. We specified the following in the methods section, “In order to calculate the inter-rater reliability and effect of experience on reliability, Cohen’s K was initially determined between the following pairs of raters: R1 and R2; R1 and R3; and R2 and R3 (monofilament and neurothesiometer) and R4 and R5 (tuning fork tests). Fleiss’ К was then calculated to determine the overall reliability between raters R1-R3 (34).” So as not to be repetitive, we have left this explanation in the statistical analysis, and kept the results as presented originally, with further elaboration available in the tables, if required.

Comment 8: Page 8, Line 24 - Would it be better to describe it as a ‘pooling’ of the data for experienced podiatrists rather than pairing (sorry maybe a bit pedantic).

Author: Thank you for your recommendation. We have changed ‘pairing’ to ‘pooling’ to satisfy your request.

Discussion
Comment 9: Page 9, Line 29 - 32 - 'Four and 10-site monofilament tests did not appear to be related to level of clinician experienced and overall appears to be reliable - I tend to agree, but wonder if you should state this since your results do show that there is greater reliability between raters with more experience (i.e. R2 and R3 K = 0.72, compared with K = 0.55) for the 4-site method.

Thank you for your suggestion. We have changed the sentence to be less ambiguous and controversial. It now reads, “Monofilament tests overall, appear to be reliable with clinical experience possibly increasing the reliability of the four-site test”.

Comment 10: Page 10, Line 5 - 8 - I think this may be overstating greater reliability of the four-site test over the 10 site test. This only applied when comparing 3 raters. When you are comparing pairs of raters you report a range from 0.55 - 0.72 which was not substantially different to figures for the 10-site method (only comparing raters 2 and 3 was substantially higher).

Author: Thank you for your recommendation. We agree and have now changed the sentence to reflect your concern. It now reads as follows, “Our study supports the relatively high inter-rater reliability previously demonstrated, with four- and 10-site tests demonstrating similar reliability overall and slightly higher reliability of the four-site test on comparison of the most experienced pairing of clinicians”.

Comment 11: Page 10, Line 7 - 10 - 'The higher intra-rater reliability previously described was not supported by our research’ - can you please specify to which test/s you are referring. It comes on the end of talking about both the 4 and 10 site monofilament testing (but is also a paper focusing on the tuning fork?)

Author: Thank you for highlighting the incorrect reference. We have since changed the reference to the correct one. The section is now clearly and consistently discussing the monofilament. The section now reads, “Lastly, level of agreement between the four- and 10-site test in 1915 people with diabetes was recently shown to be high (K: 0.87) (26) indicating that these tests may be similarly reliable. Our study supports the relatively high inter-rater reliability of the four- And 10-site 10g monofilament tests previously reported. The inter-rater reliability of four- and 10-site tests from this present study demonstrated similar levels of reliability overall, although experience improved reliability for the four-site test. The excellent intra-rater reliability previously described in the nine-site monofilament test (20) was not replicated in the four or 10 site tests used in our study. The large range of intra-rater reliability of the monofilament (fair to substantial) was not associated with greater clinical experience. As these tests rely on subjective responses from a patient, it is possible that these tests will demonstrate variability regardless of the level of experience of the clinician.

Comment 12: Page 11, Line 34 - 'Generalisable', to be consistent with other use of English (UK)

Author: Thank you, English (UK) used throughout.

References

Comment 13: There is inconsistent capitalisation with journal names. i.e. 'Archives of Internal Medicine' and 'The New England journal of medicine'

Author: Thank you for noticing the errors. We have now fixed the references to ensure consistent use of capitalisation.

Comment 14: Reference 17 - JAMA has been abbreviated rather than written in full
Author: Thank you for your suggestion. All journal names have now been written out in full.

Tables
Comment 15: Table 1 - State n=44, but report for all 50

Thank you for noticing. Now amended to ‘50’.