Westerman, T.G & Johnson, A., ***A response to the validation of the aPHQ-9 for use with Indigenous Australians***

Introduction

The adapted Patient Health Questionnaire 9 (aPHQ-9) is a translated version of the PHQ-9 for use as a depression screening tool in Indigenous men from central Australia (Brown et al., 2013). The adaptation process first involved semi-structured interviews of 22 aboriginal men of various ages from Alice Springs and surrounding regions, carried out in 2006 (Brown et al., 2012). Qualitative analysis of these interviews was then used to inform focus group discussions to select the appropriate measure to adapt (Brown et al., 2013). These focus groups were comprised of male translators, elders, or other bi-lingual Indigenous individuals from five local language groups (Brown et al., 2013). These focus groups, and their respective communities, led the choice of measure to adapt (PHQ-9) as well as the translation and back-translation process (Brown et al., 2013).

This adaptation/translation process, in and of itself, was quite thorough. However, the assumptions that the adaptation were based on, and the conclusions drawn by the authors based on the adaptation itself, raise some concerns. Firstly, the adaptation of the PHQ-9 was conducted on the assumption that no other measure of depression or depressive symptoms in Indigenous Australians was available (Brown et al., 2013). As quoted in the article itself "Screening tools for depression have not been formally validated for Aboriginal and Torres Strait Island people across multiple states and territories in Australia." (Abstract). Yet, the Westerman Aboriginal Symptoms Checklist for both adults (WASC-A) and young people (WASC-Y) had been available since 2001 and 2012 respectively. These measures were developed uniquely by an Aboriginal psychologist across rural and urban populations and with extensive consultation across different language groups. While the authors of the aPHQ-9 could have felt that the WASC-A/Y did not address their specific needs, these measures (and their development process) could have helped inform the adaptation of the aPHQ-9. This is particularly given that their own systemic review identified the WASCA as having more clinical utility than the aPHQ-9 (pg. 169: Le Grande, Thompson, 2017). Additionally, given the paucity of depression measures in Indigenous Australians, the authors of the aPHQ-9 would have had little other directly relevant literature to guide the development of a measure for mental health symptoms in this population.

In relation to this, and of greater concern, was the choice to adapt an existing measure rather than to develop a new one. The authors' made extensive reference to the problems inherent in 'transposing' western labels of symptom pathology to an indigenous population (Brown et al., 2013). However, by electing to adapt an existing (western) measure of depressive severity such as the PHQ-9, the aim then becomes finding the culturally appropriate language for assessing western constructs of depressive symptomology. This approach makes a fundamental error by incorrectly assuming that depression manifests similarly for both non-Indigenous and Indigenous Australian populations. Specifically, the Lancet published a review article discussing the factors underlying the health gap in Indigenous peoples from Australia, New Zealand, and North America (King, Smith, & Gracey, 2009). Of direct relevance to the aPHQ-9 was this statement:

"Indigenous mental health constructs are fundamentally different from those that form non-Indigenous frameworks in developed countries."

(King et al., 2009)

Therefore, the decision by the authors to translate an existing, western, measure of depression is one that ignores the growing body of literature emphasising the differences in mental health presentations between Indigenous and non-Indigenous peoples.

Most vitally, the cultural validation of the aPHQ-9 fails to address arguably the most significant contributor to test bias with Aboriginal Australians: practitioner error in the administration of psychological tests (Westerman, 2003). Whilst the authors do acknowledge the issue of training and the culturally appropriate application of the aPHQ-9 as challenges for its widespread implementation, they fail to recommend viable solutions to addressing these issues. Comparatively, tools such as the WASCA and WASCY incorporate thorough administration training in addition to clinical and cultural validation guidelines specifically developed to combat this test bias. Additionally, the development and validation of the Aboriginal Mental Health Cultural Competency Profile (Westerman, 2003; Westerman & Sheridan, in preparation) directly addresses the issue of cultural competence associated with the application of tests. (Westerman, 2003, 2010; Kearins, 1981).

The conclusions of the adaptation process are beyond the scope of what the authors had demonstrated. The adaption of the aPHQ-9 concludes with the statement: "*This paper outlines the steps taken to adapt the PHQ-9 for use across Indigenous Australian communities*." (Brown et al., 2013). However, the adaptation process was exclusively

informed by the perspectives of Indigenous men in central Australia (Brown et al., 2013). Indigenous cultures traditionally have distinct spheres of ‘mens business’ and ‘womens business’, and the experiences within these spheres are typically not shared with the other gender (Fredericks et al., 2014). By solely relying on consultations with Indigenous men in the development of this measure, the ability of the measure to capture the experiences of Indigenous women is greatly reduced. Further, it is well recognised that Indigenous men and women have different experiences and outcomes in the Australian healthcare system (Fredericks et al., 2017). By excluding Indigenous women from the consultation and adaptation process of the aPHQ-9, the measure risks further contributing to the gender divide in treatment quality and outcomes for Indigenous Australians.

Validation

To validate the aPHQ-9, 500 Indigenous individuals from ten primary health care services across Australia were recruited (The Getting it Right Collaborative Group, 2019). Participants completed the aPHQ-9, the MINI 6.0.0, and other demographic questions (The Getting it Right Collaborative Group, 2019). There are several concerns with the procedure of this validation study, and how these procedural choices risk biasing the study’s conclusions.

Firstly, the MINI has not been validated for use in Australian Indigenous populations. Consequently, it cannot be said with confidence whether the MINI is accurately identifying depressive episodes or not. In effect, the study is using an **unvalidated measure to validate another unvalidated measure**. The use of the MINI as the ‘gold-standard’ for depression diagnosis also does not seem to be consistent with the foundations that the adaptation was based on. A key concept of the adaptation was that the use of ‘western labels of symptoms’ was not appropriate for Indigenous Australians (Brown et al., 2013). However, by using the MINI (unvalidated for Indigenous Australians) as a gold standard, the validation study is effectively assessing whether the aPHQ-9 is consistent with western labels of depression and depressive episodes. This is not to say that there was a ‘better’ gold standard that the authors should have chosen, given the clear lack of depression diagnostic tools available. Rather, this is to say that the authors should not have defined ‘validation’ in terms of replicating the diagnoses provided by the MINI, when the initial purpose of the aPHQ-9 was to avoid the use of ‘western’ labels of symptom pathology.

The sampling and inclusion criteria, and their impacts on the generalisability of results also bear consideration. While the breadth of sampling (health services across Australia) is a great strength of the study, the participant recruitment within some of those services was not consistent. The authors stated that:

'At two services, staff members did not always recruit consecutive patients, sometimes selecting as potential candidates people they had met previously and believed were more likely to participate.'

The Getting it Right Collaborative Group (2019)

This statement indicates that a fifth of the recruitment was targeted and purposeful, whilst the remainder was (approximately) random. Targeted recruitment, in and of itself, is not a concern when the aims of the study involve a specific subset of the population. This study, however, purported to be representative of all Indigenous Australians, while the recruitment was targeted at individuals who *presented* as being more likely to participate (or known to the recruiters). In other words, the participation of these individuals was not determined by their own willingness to participate, but by the perceptions of the recruitment staff at a given facility. This issue of generalisability is further exacerbated by the inclusion criteria of the study. According to the study's protocol, one of the inclusion criteria was: 'Able to communicate in English sufficient to complete study instruments' (Hackett et al., 2016). In effect, this restricts participation to those individuals that could complete a clinical diagnostic interview in English. This criterion seems at odds with the aim of a measure like the aPHQ-9, which is to provide a tool for assessing Indigenous individuals in a way that is more appropriate than current 'western' measures. However, by only including those individuals that are competent in English, there is a risk of excluding those individuals who have the greatest need of a measure that can capture their depressive symptom severity.

Of greater concern, however, is the administration of these measures to individuals that have differing levels of English competence. The authors stated that measures were administered in "…English or the appropriate Aboriginal or Torres Strait Island language." (The Getting it Right Collaborative Group, 2019). This statement appears to imply the measures were translated for a given participant at the discretion of the interviewer. This is concerning for several reasons. Firstly, the purpose of a measure like the aPHQ-9 is that the wording and structure has been specifically developed to best suit the population of interest, and the aim of the subsequent validation study is to determine the extent to which this has

been achieved. However, if some individuals are receiving translated versions of the aPHQ-9, then they are receiving a measure with slightly different wording and structure. Accurately assessing the validity of this measure then becomes much more difficult, as different wordings may perform differently in assessing depressive severity. The consequences of this are further exacerbated if the translations of the aPHQ-9 were 'ad-hoc' and determined by a given translator at the time of the interview. This would mean that not only was the administered measure not consistent between language groups, it was also not consistent *within* language groups. The use of translations also implies further concerns with the use of the MINI as a 'gold standard'. If these individuals required a translator to complete the aPHQ-9, does that mean the MINI was also delivered via translator? Were this the case, this implies serious concerns about the validity results, as now the MINI being administered to each individual is also not consistent. Consequently, this would imply that different versions of the aPHQ-9 are being compared against different versions of the MINI, with little consistency within and between language groups.

Inconsistencies between language groups may also have been evident in the participants' ratings of the questionnaire itself. According to the authors, 13% of participants did not find the questions easy to understand, and 18% did not find them easy to answer. There are two likely causes of this. Firstly, the adapted wording choices for the aPHQ-9 may not have translated clearly into the respective Aboriginal or Torres Strait Islander language at the time of assessment. Difficulties in translation, if present, may have been due to consultation process in developing the aPHQ-9 only involving a central Indigenous population. While the wording choices may have been relevant and understandable for this population, it may not have translated well due to the cultural and language differences inherent in the range of Indigenous groups throughout Australia. Secondly, as noted above, the consultation process only involved groups of Indigenous men.

Beyond specific concerns with methodological choices in the approach to adaptation and validation, there is a broader concern with the apparent omission of highly relevant work in this area. As mentioned earlier, the WASC-A and WASC-Y are scales specifically developed for the measurement of depression, anxiety, impulse control, suicidal risk and cultural resilience in Indigenous adults and youth (Westerman, 2003). These measures were created, not adapted, to best capture the aspects of mental health presentation unique to Indigenous Australians. As already noted, these measures are also accompanied by comprehensive guidance in the culturally competent engagement and assessment of

Indigenous Australians which is an essential component to reduce the risk of practitioner bias in assessment and to maintain effective rapport with the client. The processes and methods undertaken in the development of the WASC-A/Y are clearly relevant to the development of a measure to assess depressive severity in Indigenous Australians. It is interesting that these measures were not discussed in the validation of the aPHQ-9, when the authors referenced a systematic review of depressive measures for Indigenous Australians which specifically mentions the utility of the WASC-A for screening for depression Le Grande, Ski, Thompson, 2017) . It is clear that both the WASC-A/Y and their developing author (Dr Tracy Westerman) could have been a highly valuable resource in the development and validation of a new measure. Yet, the WASC-A/Y was not discussed, and Dr Westerman was not consulted.

Overall, while the *aim* of the aPHQ-9 is a step in the direction of culturally appropriate mental health assessment tools, the realisation of that aim is limited and poorly generalisable. By failing to consider the work of established culturally appropriate mental health assessments (i.e. the WASC-A and WASC-Y), the authors have ignored crucial empirical evidence in the area of indigenous mental health assessment. Further, the highly limited community consultation process, precludes the effective use of the aPHQ-9 measure with Indigenous women. Finally, the validation process was strongly at odds with the aims of the adapted measure. Where the aPHQ-9 was developed to avoid the use of Western labels of symptom pathology, the validation study assessed how well the aPHQ-9 could replicate those labels in the Indigenous population. The effective and culturally appropriate measurement of mental health difficulties in Indigenous Australians is a key step towards to 'closing the gap' for this population, and future efforts in that direction are much needed and welcomed. However, that step needs to be taken in a manner that includes all Indigenous Australians and takes advantage of the increasing amount of research in the area, if it is to be taken at all.

References

Brown, A., Mentha, R., Rowley, K. G., Skinner, T., Davy, C., & O'Dea, K. (2013). Depression in Aboriginal men in central Australia: adaptation of the Patient Health Questionnaire 9. *BMC Psychiatry, 13*(1), 271. Retrieved from https://doi.org/10.1186/1471-244X-13-271. doi:10.1186/1471-244x-13-271

Brown, A., Scales, U., Beever, W., Rickards, B., Rowley, K., & O'Dea, K. (2012). Exploring the expression of depression and distress in aboriginal men in central Australia: a qualitative study. *BMC Psychiatry, 12*(1), 97. Retrieved from https://doi.org/10.1186/1471-244X-12-97. doi:10.1186/1471-244x-12-97

Fredericks, B., Clapham, K., Bainbridge, R., Collard, L., Adams, M., Bessarab, D., . . . Daniels, C. (2014). 'Ngulluck Katitj Wah Koorl Koorliny/ Us mob going along learning to research together': Drawing on Action Research to develop a literature review on Indigenous gendered health and wellbeing. *Action Learning and Action Research Journal, 20*(2), 89-113.

Fredericks, B., Daniels, C., Judd, J., Bainbridge, R., Clapham, K., Longbottom, M., . . . Ball, R. (2017). Gendered Indigenous health and wellbeing within the Australian health system: A review of the literature. Retrieved from https://eprints.qut.edu.au/115966/

Le Grande M, Ski CF, Thompson DR. Social and emotional wellbeing assessment instruments for use with Indigenous Australians: a critical review. Soc Sci Med 2017; 187: 164–173.

Hackett, M., Cass, A., Glozier, N., Skinner, T., Teixeira-Pinto, A., Askew, D., . . . Brown, A. (2016). The validation of a culturally-specific measure to identify depression in Aboriginal and Torres Strait Islander people with or without chronic disease. Retrieved from https://rsph.anu.edu.au/files/Validation-Study-Summary-Report.pdf

Kearins, J. M. (1981). Visual Spatial Memory in Australian Aboriginal children of the desert regions. Cognitive Psychology, 13, 434-460.

King, M., Smith, A., & Gracey, M. (2009). Indigenous health part 2: The underlying causes of the health gap. *The Lancet, 374*(9683), 76-85. Retrieved from http://www.sciencedirect.com/science/article/pii/S0140673609608278. doi:https://doi.org/10.1016/S0140-6736(09)60827-8

The Getting it Right Collaborative Group. (2019). Getting it Right: validating a culturally specific screening tool for depression (aPHQ-9) in Aboriginal and Torres Strait Islander Australians. *Medical Journal of Australia, 211*(1), 24-30. Retrieved from https://onlinelibrary.wiley.com/doi/abs/10.5694/mja2.50212. doi:10.5694/mja2.50212

Westerman, T. G. (2004). Guest Editorial: Engagement of Indigenous Clients in Mental Health Services: what role do cultural differences play? Australian e-Journal for the Advancement of Mental Health, 3(3).

Westerman, T.G. (2010). Engaging Australian Aboriginal youth in mental health services.

Australian Psychologist, 45, 212-222. doi: /10.1080/00050060903451790