PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Brief assessment of male depression in clinical care: Validation of the Male Depression Risk Scale Short Form in a cross-sectional study of Australian men
AUTHORS	Herreen, Danielle; Rice, Simon; Zajac, Ian

VERSION 1 – REVIEW

REVIEWER	Islam, Fakir Swinburne University of Technology, Statistics, Data Science and Epidemiology
REVIEW RETURNED	14-Oct-2021

GENERAL COMMENTS	This is a very well written paper. However, it needs some improvements and clarification. Abstract 1. This needs to give additional information on how the subset was selected for this scale validation. Please see some recent publications on how they took some subsamples to show the robustness of the findings. If robustness is checked, it will give the answer of selecting subsamples. Uddin N. Islam FMA. 2020.Psychometric evaluation of the modified 19-item Bengali version of WHOQOL scale using Rasch analysis: a cross-sectional study of a rural district in Bangladesh. 2020. BMC Psychology, Vol 8, No.1. Article # 44. Uddin N. Islam FMA. 2020. Psychometric evaluation of the modified Kessler seven-item version (K7) for measuring psychological distress using Rasch analysis: a cross-sectional study in a rural district of Bangladesh. BMJ Open, Vol 10, No.2. e034523. Uddin N. Islam FMA and Mahmud AA. Psychometric evaluation of an interview-administered version of the Kessler 10-item questionnaire (K10) for measuring psychological distress in rural Bangladesh. BMJ Open. 2018. http://hdl.handle.net/1959.3/444034. [30%, Q2] 1. In the abstract, some scales have been mentioned as the primary and secondary outcomes. However, in the results section, only MDRS-7 is said. Would you please give adequate information about the other scales? Uddin and Islam proposed K7 from K10. This is an opportunity to check that double validate tool. 2. Table 2: Variance explained 38.80% (overall) for Time 1; what is this value for Time 2. I prefer to show the summary information at Time 2 as it is done in Time 1.

	3. Table 3: Presenting the number (%) and the odds ratio in the same column has made it difficult to understand. Would you please insert different columns for N(%)? Also, n(%) is given in the first column; then, it is not necessary to give in the subsequent columns, i.e., 11 (6) instead of 11 (6%). 4. Are these models adjusted? If these are unadjusted, please present the adjusted models. 5. There are a number of factor variables, e.g., education, income, job status. Where are these variables used? A summary table could be given to show if there were any differences in scales or models for different factors.

REVIEWER	Genuchi, Matthew Boise State University
REVIEW RETURNED	22-Nov-2021

GENERAL COMMENTS	Thank you very much for the opportunity to review BMJ open 2021-053650. The authors recruited a community sample of younger and older Australian men in order to examine a short form of the MDRS-22. I am quite familiar with the MDRS-22, and it is the most extensively studied externalizing symptom scale available to examine atypical depression in men. The authors studied a shorter version of the MDRS-22 that could maintain the strong psychometric properties of the original scale but also be clinically valuable in primary care settings. I agree with the authors regarding the importance and potential impact of their work. Increased use of externalizing symptom measures in primary care could have significant impacts on the recognition of depression and suicide in Australian men, as well as men in other western countries with high male suicide rates (e.g. United States). Furthermore, a brief measure would be much more simple to use in primary care settings or other settings where brief symptom screenings are necessary or beneficial (e.g. outpatient mental health care, pain clinics, etc.). They propose some interesting findings regarding how the MDRS-7 was indicative of depressive symptoms over time (e.g. moderate MDRS-7 scores were associated with later depression) and that the MDRS-7 increased the identification of individuals experiencing current suicidality, compared to prototypic symptoms alone. I believe that the authors' work would be of interest to men's health researchers, such as me, but likely to others who are interested in the intersection of men's health and primary care. Therefore I believe the potential impact of their work is quite broad. I believe that the authors methods and analyses are sound, yet I do have one question in this area. The authors note (e.g. p. 13, line 348) that two items on the MDRS-7 loaded weakly on the full MDRS-7 scale, and the authors provide a rationale for this finding (this is true the Cronbach's alpha is quite a bit lower for these two items). The authors then appear to note that they retained these items for conceptual reasons. Are the authors stating that under different circumstances (for example, just looking at the reliability statistics) they would not retain these items in the scale, but in this case, they decided it was conceptually important? If that's the case, is there existing literature that supports the authors' rationale? It would help if they could explain a bit further here. I am not a scale development expert, so this may be standard practice

	(and I am unaware), but it might be clearer if the authors can reference if/how this practice is standard and why. Thanks for the opportunity to review this manuscript and I wish the authors the best in their continued efforts.

REVIEWER	Macdonald, Jacqui Deakin University
REVIEW RETURNED	26-Nov-2021

GENERAL COMMENTS	This paper reports on development and validation of a 7-item short-form of the Male Depression Risk Scale. Current instruments may fail to detect common male presentations of mental health problems and so development of measures that capture appropriate symptoms is an extremely important endeavour. The rationale for this instrument and the problem the study addresses are well made in a well written introduction. Abstract: I found the abstract difficult to understand until I had read the manuscript, so I would suggest it would benefit from some rewording to clarify the keys constructs measured and methodologies used. In short, it might benefit from a stronger narrative of what was done. In the Setting section, it seems imprecise to say participants were recruited via an online survey. In the Participants section, I do not think it is necessary to include 'completed measure of externalising and prototypic depression symptoms' but the word prototypic should be included in the measures section. I have a concern about describing the MDRS as a measure of externalising given that is only part of what it captures. I found this confused me when I read the abstract and the paper generally because the MDRS was described as a scale of 'externalising and male specific depression symptoms' but was treated conceptually only as a measure of externalising with the latter term used interchangeably with the measure. The Abstract Results section loses clarity because the first analysis reported is not a typical examination of construct validation. Rather it reports on prevalence of symptom groupings and this was unexpected. Given this is not a representative sample, it is important to be very clear that these are sample specific proportions. In the abstract conclusion, would it not be more precise to describe the MDRS-7 as a tool that includes externalised symptoms rather than one that sounds like it only measures externalised symptoms? Also, I would argue that the short-form is not ready to be described as an instrument that may be used in clinical settings. I recommend a more cautious approach suggesting that this is a first and promising step towards development of an instrument that should be further assessed for clinical utility. Introduction: I think it would help to provide more explicit aims that align directly to each of the analyses. I was surprised when I realised that the MDRS was to be paired with the PHQ-9 to create symptom groupings. I think this is a novel and appropriate way to investigate the measure but the setting up of this as a research question was a little opaque. At the moment, it is left to the reader to join the dots that the mention of 'mental illness subgroups' relates to those found in the previous research on

	prototypic v others. All the information is there but it could be signposted with more clarity. An option might be to make a statement prior to the aims that if a short form is to have clinical utility it would need to be able to identify risk across the male depression presentations as per the prior findings, such that prototypic, mixed and male-specific symptom patterns could be identified. Therefore, aim …… Methods: Line 195: typo - thar should be that. Information on who was excluded in the final sample seemed out of place here. I would have expected to see this in the Results not the statistical analyses section. Is there any demographic information to indicate whether those who were included differed from those who were not? Line 233: I suggest changing 'demonstrate' to 'investigate' or 'explore'. Can the authors elaborate on what they mean by the scores corresponding to cut-off percentiles in the MDRS-22? Perhaps further information on this can be given in the supplementary materials. Line 237: When stating that individuals were classified into depression groups, can the authors explicitly state that they are those identified in the prior study. Line 240 (and elsewhere), as previous mentioned, I am concerned that the MDRS > 5 score is being described as externalising. First not all symptoms assess externalising and second a participant could theoretically answer zero on all externalising items and answer 'most of the time' on 'I bottled up my negative feelings' and 'I had unexplained aches and pains' for a score of 6. This can be resolved by simply broadening the labelling of the construct. Line 225. States EFA with ML estimation was used and the table 2 states that a principal components analysis was used. Consider reworking the statistical analyses section so that it is written with an active voice: https://www.bmj.com/about-bmj/resources-authors/house-style instead of for example '…the k10 was used…'. BMJ encourages use of the first person to avoid passive writing. Results: Can the authors provide population level comparison information so that readers can ascertain the nature and extent of biases in the sample? And also sample differences from original to retained participants at time 2? Line 272. Is the word sensitivity technically correct? Line 295. 'effect of age' sounds causal – these are age group differences Line 314. No need for the word 'undoubtedly'.

	Discussion: Overall, my main recommendation is to tone down the implications of the findings. This is the first look at the validity and potential utility of a short form. The results suggest it has promise but much more work is needed to further assess its value across samples. This includes full assessment of predictive psychometrics in a sample that also conducts assessments with clinical diagnostic instruments. It would be possible then to also note that while this is needed, there is a circular problem of under diagnosis of men with these so-called 'gold standard' assessments. Line 353… where there is comment on the lower loadings for the alcohol and drug use items, the authors argue for the retention of these items because they are markers of depression and suicidality but there is also an argument that co-morbid factors are not appropriate as joint indicators. i.e., it is like using anxiety to measure depression. They will be related but they are not indicators of the same condition. In the latter example, in combination they indicate a broader construct of psychological distress. Including substance use in the male depression instrument might be muddying the construct under investigation. I think there could be more nuance in discussion around this point. See Macdonald et al. (2020) for more on this argument: https://www.frontiersin.org/articles/10.3389/fpsyt.2020.578114/full Once some comparison of the sample is made with population characteristics, limitations should further address the sampling biases There is mention of the online sample reducing generalisability. There is some interesting literature on sampling biases from online recruitment that would add value to the point here. A limitation also worth noting is that risk might not look the same for older and younger participants and so selecting the same items across age groups might not be appropriate. For example, unexplained aches and pains are common in older age and so understanding their age-related link to depression is important if this indicator is to be used for the older men. This might be akin to exclusion of somatic symptoms in perinatal mothers because they may indicate expected physiological responses to pregnancy and birth. In this study the sample was also divided by age 65. There might be other developmental considerations in mid adulthood that mean findings would be different again at that life stage. However, there are benefits to a measure that can assess at different points of the lifespan. Can the authors discuss this in the context of their measure? Line 441, a suggested change would be to replace the word 'demonstrates' with 'provides emerging support for…'

VERSION 1 – AUTHOR RESPONSE

REVIEWER 1:
ABSTRACT
1. This needs to give additional information on how the subset was selected for this scale validation. Please see some recent publications on how they took some subsamples to show the robustness of the
findings. If robustness is checked, it will give the answer of selecting subsamples.
Uddin N. Islam FMA. 2020.Psychometric evaluation of the modified 19-item Bengali version of WHOQOL
scale using Rasch analysis: a cross-sectional study of a rural district in Bangladesh. 2020. BMC Psychology, Vol 8, No.1. Article # 44.
Uddin N. Islam FMA. 2020. Psychometric evaluation of the modified Kessler seven-item version (K7) for
measuring psychological distress using Rasch analysis: a cross-sectional study in a rural district of Bangladesh. BMJ Open, Vol 10, No.2. e034523.
Uddin N. Islam FMA and Mahmud AA. Psychometric evaluation of an interview-administered version of
the Kessler 10-item questionnaire (K10) for measuring psychological distress in rural Bangladesh. BMJ
Open. 2018. http://hdl.handle.net/1959.3/444034. [30%, Q2]
We thank the reviewer for this suggestion. Various approaches exist for scale reduction. We acknowledge
the references provided for the Uddin et al. papers, where Rasch analysis was used to explore scale validity. Whilst we value the suggestion to consider alternative approaches, our approach differs from Rasch analysis, and instead utilised confirmatory factor analysis (a widely accepted psychometric method).
We believe that extension of the current analyses would not significantly improve the current manuscript
findings. However, we have included the second reference in relation to Rasch models on page 8, lines
280-281.
2. In the abstract, some scales have been mentioned as the primary and secondary outcomes. However,
in the results section, only MDRS-7 is said. Would you please give adequate information about the other
scales? Uddin and Islam proposed K7 from K10. This is an opportunity to check that double validate tool.
We appreciate the reviewer's suggestion to validate the K7. However, we do not believe that addressing
this is within the scope of the current study. Furthermore, adding the necessary information would exceed
the word limit. In the results section, we describe results relating to mental illness and suicidality, but have
now indicated in brackets that mental illness refers to the K10, and suicidality refers to the PHQ-9 for clarity. Please see page 2, lines 56-57.
3. Table 2: Variance explained 38.80% (overall) for Time 1; what is this value for Time 2. I prefer to show
the summary information at Time 2 as it is done in Time 1.
We appreciate the reviewer's request to add this. However, we would like to clarify that we cannot do this

because item loadings at Time 2 were derived using confirmatory factor analysis as specified in text and in
the footer of Table 2. This statistic (% variance explained) is not derived during confirmatory modelling.
3
4. Table 3: Presenting the number (%) and the odds ratio in the same column has made it difficult to understand. Would you please insert different columns for N(%)? Also, n(%) is given in the first column;
then, it is not necessary to give in the subsequent columns, i.e., 11 (6) instead of 11 (6%).
We thank the reviewer for making this suggestion. We have now adjusted the presentation of Table 3 to
improve readability. We have also removed the % symbol from subsequent columns.
5. Are these models adjusted? If these are unadjusted, please present the adjusted models.
The models presented in Table 3 are fully adjusted for previous diagnosis of depression. We have now
revised OR to AOR in the table headings. For brevity, we have not included the OR of previous depression
for fear of further complicating the table.
6. There are a number of factor variables, e.g., education, income, job status. Where are these variables
used? A summary table could be given to show if there were any differences in scales or models for different factors.
Given the aim of the paper is primarily focussed on deriving a short form of the MDRS, it has been necessary to present refined analyses so as not to detract from the overall aim, nor exceed the word length. We have added the suggestion for future research to examine measurement invariance according
to variables such as education level, income, and cultural background. Please see page 14, lines 542-543.
REVIEWER 2:
1. The authors note (e.g. p. 13, line 348) that two items on the MDRS-7 loaded weakly on the full MDRS7 scale, and the authors provide a rationale for this finding (this is true the Cronbach's alpha is quite a bit
lower for these two items). The authors then appear to note that they retained these items for conceptual reasons. Are the authors stating that under different circumstances (for example, just looking
at the reliability statistics) they would not retain these items in the scale, but in this case, they decided it
was conceptually important? If that's the case, is there existing literature that supports the authors' rationale? It would help if they could explain a bit further here. I am not a scale development expert, so
this may be standard practice (and I am unaware), but it might be clearer if the authors can reference if/how this practice is standard and why.
We thank the reviewer for their positive feedback and helpful comments. The two items (e.g., Using drugs
provided temporary relief and I needed alcohol to help me unwind) loaded at .48 and .42 in the overall sample. These loadings actually exceed the minimum recommended factor loading of .32 (DeVellis, 2016).
Although some authors do argue that .32 is relatively low (e.g., Comrey & Lee, 2013), this remains the recommended threshold from which item loadings can be interpreted (Tabachnick & Fidell, 2007). The two
lower loading items we observed (.42 and .48) exceeded the .32 loading threshold, as well as Cohen's .30

threshold to indicate a moderate correlation. We have also clarified in the revised manuscript that the decision to retain these items was both statistical (e.g., item loadings > .32) and conceptual (e.g., in order
4
for the brief version of the MDRS-7 to represent at least one item from each of the 6 domains assessed by
the original MDRS-22). Please see pages 12-13, lines 564-576.
REVIEWER 3:
ABSTRACT
1. I found the abstract difficult to understand until I had read the manuscript, so I would suggest it would
benefit from some rewording to clarify the keys constructs measured and methodologies used. In short,
it might benefit from a stronger narrative of what was done.
We have made some edits to the abstract, including additional detail about what was done. Given structural and word length requirements, we are unable to provide further detail here, but hope these changes have provided additional clarity.
2. In the Setting section, it seems imprecise to say participants were recruited via an online survey.
We have removed reference to via an online survey (please see page 2, line 40).
3. In the Participants section, I do not think it is necessary to include 'completed measure of externalising and prototypic depression symptoms' but the word prototypic should be included in the measures section.
Thank you for this suggestion. We have now adjusted the wording for this section (please see page 2, lines
47-52).
4. I have a concern about describing the MDRS as a measure of externalising given that is only part of
what it captures. I found this confused me when I read the abstract and the paper generally because the
MDRS was described as a scale of 'externalising and male specific depression symptoms' but was treated
conceptually only as a measure of externalising with the latter term used interchangeably with the measure.
We thank the reviewer for this suggestion. We agree that there is some conceptual complexity regarding
terminology as not all items in the MDRS-22 reflect externalising symptoms (e.g., items assessing emotion
suppression such as I bottled up my negative feelings and somatic symptoms such as I had unexplained
aches and pains). We initially adopted this terminology to be consistent with existing literature. We have,
however, made adjustments to wording throughout the manuscript to consistently (and more accurately)
reflect that the MDRS is a measure of 'externalising and male-type symptoms' associated with major depression in men.
5. The Abstract Results section loses clarity because the first analysis reported is not a typical examination of construct validation. Rather it reports on prevalence of symptom groupings and this was
unexpected. Given this is not a representative sample, it is important to be very clear that these are sample specific proportions.
5

We thank the reviewer for this comment. In addressing previous comments to provide more clarity about
what was done, we have had to remove this section due to word length requirements.
6. In the abstract conclusion, would it not be more precise to describe the MDRS-7 as a tool that includes
externalised symptoms rather than one that sounds like it only measures externalised symptoms?
Thank you for this suggestion. We have now amended the wording to refer to a measure that includes externalised and male-type symptoms. Please see page 2, line 62.
7. Also, I would argue that the short-form is not ready to be described as an instrument that may be used in clinical settings. I recommend a more cautious approach suggesting that this is a first and promising step towards development of an instrument that should be further assessed for clinical utility.
Thank you for this suggestion. We have now added a sentence indicating that future research is needed to
further examine the clinical utility of this tool and have replaced the word 'use' with 'field trials'. Please see
page 2, line 63.
INTRODUCTION
8. I think it would help to provide more explicit aims that align directly to each of the analyses.
I was surprised when I realised that the MDRS was to be paired with the PHQ-9 to create symptom groupings. I think this is a novel and appropriate way to investigate the measure but the setting up of this as a research question was a little opaque. At the moment, it is left to the reader to join the dots that the mention of 'mental illness subgroups' relates to those found in the previous research on prototypic v others. All the information is there but it could be signposted with more clarity.
An option might be to make a statement prior to the aims that if a short form is to have clinical utility it would need to be able to identify risk across the male depression presentations as per the prior findings,
such that prototypic, mixed and male-specific symptom patterns could be identified. Therefore, aim ……
We thank the reviewer for this suggestion. We have now provided additional information in the introduction to clarify the rationale for the study aims on page 5, lines 173-177.
METHODS
9. Line 195: typo - thar should be that.
Thank you for noticing this. We have now corrected this (please see page 7, line 254.
10. Information on who was excluded in the final sample seemed out of place here. I would have expected to see this in the Results not the statistical analyses section. Is there any demographic information to indicate whether those who were included differed from those who were not?
6
We thank the reviewer for this comment. We have now moved this information to a separate paragraph
under 'Analytic sample' (please see pages 7-8, lines 269-275). We have also provided demographic information on the final sample in Table 1.
11. Line 233: I suggest changing 'demonstrate' to 'investigate' or 'explore'.
Thank you for this suggestion. We agree that investigate is a better fit and have changed this (please see
page 8, line 296).
12. Can the authors elaborate on what they mean by the scores corresponding to cut-off percentiles in
the MDRS-22? Perhaps further information on this can be given in the supplementary materials.
Thank you for this suggestion. We have clarified the cut-off percentiles on page 8, lines 298-302.
13. Line 237: When stating that individuals were classified into depression groups, can the authors explicitly state that they are those identified in the prior study.

We have now clarified that these depression groups were based on a previous study and have included the
reference on page 9, lines 330-334.
14. Line 240 (and elsewhere), as previous mentioned, I am concerned that the MDRS > 5 score is being
described as externalising. First not all symptoms assess externalising and second a participant could theoretically answer zero on all externalising items and answer 'most of the time' on 'I bottled up my negative feelings' and 'I had unexplained aches and pains' for a score of 6. This can be resolved by simply
broadening the labelling of the construct.
We thank the reviewer for this comment. As per our response to comment 4, we have now replaced 'externalising' with 'externalising and male-type' throughout the document.
15. Line 225. States EFA with ML estimation was used and the table 2 states that a principal components
analysis was used.
Thank you for noticing this and altering us to this error. We can confirm that EFA with ML was used and
have updated the table accordingly to reflect this.
16. Consider reworking the statistical analyses section so that it is written with an active voice: https://www.bmj.com/about-bmj/resources-authors/house-style instead of for example '…the k10 was used…'. BMJ encourages use of the first person to avoid passive writing.
Thank you for this suggestion. We have now made some changes to the statistical analyses section as
requested. Please see page 8-9, lines 277-344.
7
RESULTS
17. Can the authors provide population level comparison information so that readers can ascertain the nature and extent of biases in the sample? And also sample differences from original to retained participants at time 2?
Thank you for this suggestion. We have now added in some information comparing the sample with 2016
Australian Census data. We have also added Time 2 sample characteristics to Table 1. Please see pages 9-
10, lines 355-376.
18. Line 272. Is the word sensitivity technically correct?
The word sensitivity here has been removed. We have also removed reference to 'sensitivity' throughout
the manuscript to avoid any confusion.
19. Line 295. 'effect of age' sounds causal – these are age group differences
This has been changed to 'age group differences'. Please see page 8, line 303.
20. Line 314. No need for the word 'undoubtedly'.
This word has now been removed (please see page 11, line 485).
DISCUSSION
21. Overall, my main recommendation is to tone down the implications of the findings. This is the first look at the validity and potential utility of a short form. The results suggest it has promise but much more work is needed to further assess its value across samples. This includes full assessment of predictive psychometrics in a sample that also conducts assessments with clinical diagnostic instruments. It would be possible then to also note that while this is needed, there is a circular problem
of under diagnosis of men with these so-called 'gold standard' assessments.
We thank the reviewer for this suggestion. We have toned down the implications of the findings throughout the discussion section.

22. Line 353… where there is comment on the lower loadings for the alcohol and drug use items, the authors argue for the retention of these items because they are markers of depression and suicidality but there is also an argument that co-morbid factors are not appropriate as joint indicators. i.e., it is like
using anxiety to measure depression. They will be related but they are not indicators of the same condition. In the latter example, in combination they indicate a broader construct of psychological distress. Including substance use in the male depression instrument might be muddying the construct under investigation. I think there could be more nuance in discussion around this point. See Macdonald
et al. (2020) for more on this argument:
8
We thank the reviewer for this comment. Since repeating our analyses, the two items now load moderately. Importantly, all items loaded above the minimum recommended factor loading of .32 (DeVellis, 2016). As mentioned in the manuscript, the decision to retain these items was both statistical
(e.g., item loadings > .32) and conceptual (e.g., in order for the brief version of the MDRS-7 to represent at
least one item from each of the 6 domains assessed by the original MDRS-22).
As discussed in our paper, substance use has been shown to be a core feature of men's experiences of
depression (Addis, 2008; Coleman, 2015). We recognise that alcohol and drug use may reflect a comorbidity (Macdonald et al., 2020), or alternatively reflect maladaptive coping (Cavanagh, Wilson, Kavanagh, & Caputi, 2017). These are important questions for future research to explore. We nonetheless
believe that the MDRS-7 is an important tool to assist with furthering our understanding of men's experiences of psychological distress and suicide risk.
We have now provided some additional information in our discussion section and have included the Macdonald et al. (2020) and Cavanagh et al. (2017) reference (please see pages 13, lines 568-576).
Addis, M. E. (2008). Gender and depression in men. Clinical Psychology: Science & Practice, 15(3), 153-168.
doi:10.1111/j.1468-2850.2008.00125.x
Cavanagh, A., Wilson, C., J., Kavanagh, D., J., & Caputi, P. (2017). Differences in the expression of symptoms
in men versus women with depression: A systematic review and meta-analysis. Harv Rev Psychiatry, 25(1), 29-38. doi:10.1097/hrp.0000000000000128
Coleman, D. (2015). Traditional masculinity as a risk factor for suicidal ideation: Cross-sectional and prospective evidence from a study of young adults. Archives of Suicide Research, 19(3), 366-384. doi:10.1080/13811118.2014.957453
Comrey, A. L., & Lee, H. B. (2013). A first course in factor analysis (P. Press Ed. 2nd ed.). Hillsdale, N.J: L.
Erlbaum Associates.
DeVellis, R. F. (2016). Scale development: Theory and applications (4th ed. Vol. 26). Thousand Oaks, Calif:
Sage Publications.
Macdonald, J. A., Greenwood, C. J., Francis, L. M., Harrison, T. R., Graeme, L. G., Youssef, G. J., . . . Olsson,
C. A. (2020). Profiles of Depressive Symptoms and Anger in Men: Associations With Postpartum Family Functioning. Frontiers in psychiatry, 11, 578114-578114. doi:10.3389/fpsyt.2020.578114
Tabachnick, B. G., & Fidell, L. S. (2007). Using multivariate statistics (5th ed.). Boston, MA: Pearson/Allyn &
Bacon.
23. Once some comparison of the sample is made with population characteristics, limitations should

further address the sampling biases
Thank you for this suggestion. We have now added in some information comparing the sample with 2016
Australian Census data and have acknowledged this in our limitations section. Please see pages 9-10, lines
355-376.
24. There is mention of the online sample reducing generalisability. There is some interesting literature
on sampling biases from online recruitment that would add value to the point here.
Thank you for this suggestion. We have included the below reference to support this comment in our discussion section. Please see page 14, line 542.
9
Choi, I., Milne, D. N., Glozier, N., Peters, D., Harvey, S. B., & Calvo, R. A. (2017). Using different Facebook
advertisements to recruit men for an online mental health study: Engagement and selection bias. Internet
Interventions, 8, 27-34. https://doi.org/https://doi.org/10.1016/j.invent.2017.02.002
25. A limitation also worth noting is that risk might not look the same for older and younger participants
and so selecting the same items across age groups might not be appropriate. For example, unexplained
aches and pains are common in older age and so understanding their age-related link to depression is
important if this indicator is to be used for the older men. This might be akin to exclusion of somatic symptoms in perinatal mothers because they may indicate expected physiological responses to pregnancy and birth. In this study the sample was also divided by age 65. There might be other developmental considerations in mid adulthood that mean findings would be different again at that life stage. However, there are benefits to a measure that can assess at different points of the lifespan. Can
the authors discuss this in the context of their measure?
We thank the reviewer for this comment. One of the strengths of the MDRS-22 is that the items are designed to gauge the individual's perception of their own recent behaviour. Specifically, the risk-taking
item states 'I took unnecessary risks', rather than listing specific behaviours. Thus, although the actual risktaking behaviours older men are engaging in might be different to younger men, the wording of the items
is broad enough to be applicable to various ages. For example, risk-taking behaviours for older males may
include ignoring medical advice or driving with impaired vision, whilst younger males may be more likely to
engage in risky behaviours such as driving at high speeds or while under the influence of alcohol or other
drugs.
Importantly, items retained in the MDRS-7 were those that performed best in both the younger and older
sample. We have previously explored whether the MDRS-22 is suitable for use with older males, with results demonstrating that the MDRS-22 is an appropriate measure of externalising and male-typical symptoms across the lifespan, including males aged 65+ years (Herreen et al., 2021). Whilst these symptoms may be less prevalent in older compared to younger males, they still appear to be important
indicators of psychological distress and suicide risk across the lifespan. We have added additional

information in the discussion section acknowledging that the MDRS-7 items were those that performed
best in both younger and older males and suggested that future research should examine the psychometric
properties of the MDRS-7 in clinical samples of men across the lifespan (please see page 15, lines 549-551).
Herreen, D., Rice, S., Ward, L., & Zajac, I. (2021). Extending the Male Depression Risk Scale for use with
older men: The effect of age on factor structure and associations with psychological distress and history of
depression. Aging & Mental Health, 1-9. https://doi.org/10.1080/13607863.2021.1947966
26. Line 441, a suggested change would be to replace the word 'demonstrates' with 'provides emerging
support for…'
Thank you for this suggestion. We have made this amendment (please see page 15, lines 561-562).

VERSION 2 – REVIEW

REVIEWER	Islam, Fakir Swinburne University of Technology, Statistics, Data Science and Epidemiology
REVIEW RETURNED	06-Feb-2022

GENERAL COMMENTS	congratulations!!

REVIEWER	Genuchi, Matthew Boise State University
REVIEW RETURNED	13-Jan-2022

GENERAL COMMENTS	Thanks very much to the authors for their willingness to address my concerns in their revision. I believe they have fully addressed my comments.

REVIEWER	Macdonald, Jacqui Deakin University
REVIEW RETURNED	13-Feb-2022

GENERAL COMMENTS	I thank the authors for their considered revisions that addressed my feedback. I believe this paper will make an important contribution to the literature.