Investigating age-related differences in responses to the screening items for internalising disorders in three national surveys

Author:
Buchan, Heather; Sunderland, Matthew; Carragher, Natacha; Batterham, Philip; Slade, Tim

Publication details:
Journal of Affective Disorders
v. 152-154
pp. 229-236
0165-0327 (ISSN)

Publication Date:
2014

Publisher DOI:
http://dx.doi.org/10.1016/j.jad.2013.09.015

Investigating age-related differences in responses to screening items for internalising disorders in three national surveys

Heather Buchan[1]

Matthew Sunderland[1]

Natacha Carragher[1]

Philip Batterham[2]

Tim Slade[1]

Manuscript in preparation for Journal of Affective Disorders

[1]NHMRC Centre of Research Excellence in Mental Health and Substance Use, National Drug and Alcohol Research Centre, University of New South Wales, Sydney, New South Wales, Australia

[2]Centre for Mental Health Research, the Australian National University, Canberra, Australian Capital Territory, Australia

Corresponding author
Heather Buchan
National Drug and Alcohol Research Centre
University of New South Wales,
Sydney NSW 2052
Australia

P: +61 (2) 8936 1010
F: +61 (2) 9385 0222
E: h.buchan@unsw.edu.au

Abstract

Background: Epidemiological studies typically report lower prevalence of mental disorders among older adults relative to middle-aged and young adults. A possible explanation is that age-related bias in the screening items of diagnostic instruments leads to older adults being differentially screened out of the full assessment. This study investigated potential age-related bias in screening items for internalising disorders in three epidemiological surveys.

Method: Measurement invariance was estimated for the internalising disorder screening items in the 2007 and 1997 Australian National Survey of Mental Health and Wellbeing, and the 2001-2002 National Epidemiologic Survey on Alcohol and Related Conditions. These surveys assessed mental disorders using the Composite International Diagnostic Interview (CIDI) and the Alcohol Use Disorder and Associated Disabilities Interview Schedule–DSM-IV Version (AUDADIS-IV). A series of multi-group confirmatory factor analyses (CFA) were performed for each survey across older (65-85 years), middle (35-64 years) and young (16-34 years) adults.

Results: Differences between successive CFA models for each survey were negligible, indicating measurement invariance across age groups for the CIDI and AUDADIS-IV screening items.

Limitations: The number of items and symptoms representing internalising pathology differed between surveys. The samples excluded people in aged-care institutions.

Conclusions: While findings do not rule out that other measurement errors may be present (e.g., age-related bias in the remaining items), these results support the validity of the screening items in the CIDI and AUDADIS-IV. Low prevalence estimates of internalising disorders in older adults are unlikely attributable to differences in response patterns to screening items.

Keywords: old age, mental disorders, screening, CIDI, AUDADIS-IV, internalising

1. Introduction

Mental disorders in old age are associated with decreased quality of life and physical health, and increased psychological distress and mortality (Beekman et al., 1998; Blazer, 2003; Doraiswamy et al., 2002; Lenze et al., 2001; Schulz et al., 2002; Wetherell et al., 2004). With an aging population across the world (United Nations, 2002), determining accurate prevalence rates of mental disorders in old age is increasingly important for optimising health care services.

Internalising disorders reflect a propensity to experience distress inwards (i.e., unipolar mood and anxiety disorders) and epidemiological studies have provided inconsistent estimates of prevalence rates of these disorders in older adults (Riedel-Heller et al., 2006; Volkert et al., 2013). However, one finding that is commonly reported is that prevalence is significantly lower in older adults than in middle and younger age groups (Jorm, 2000; Kessler et al., 2010a; Kessler et al., 2010b; McEvoy et al., 2011; Trollor et al., 2007).

There is considerable debate in the literature over whether lower prevalence of internalising disorders in old age reflects true effects in the population or methodological errors. Symptoms of internalising disorders may in fact decline in old age due to protective psychological factors that develop as people age, such as increased emotional control, better adaptive coping strategies, and decreased emotional responsiveness (Blazer and Hybels, 2005; Ernst and Angst, 1995; Henderson et al., 1998; Jorm, 2000). Conversely, diagnostic criteria and the manner in which they are operationalised in fully structured lay-administered instruments may be biased against older adults, leading to inaccurate case ascertainment in epidemiological studies (Gallo et al., 1994; Henderson et al., 1998; O'Connor and Parslow, 2009).

Indeed, a study of the Composite International Diagnostic Interview (CIDI) depression screening items in an Australian national survey found that older adults (65+

years) were significantly less likely to endorse the screening items than younger adults (O'Connor and Parslow, 2010a). Further, these age group differences were larger for the screening items than for diagnostic items in the full assessment, which tend to be shorter and simpler. O'Connor and Parslow (2010a) suggest that the lower rate of item endorsement by older adults may be due to the relatively high cognitive demand required by the CIDI screening items. Respondents who have symptoms who can manage the screening questions may then also be able to respond accurately to the full diagnostic questions. This is supported by a previous study that compared responses to the CIDI screening items with the Kessler Psychological Distress Scale (K-10; Kessler et al., 2002), which has shorter and simpler items. It was found that disagreements between corresponding CIDI and K-10 items significantly increased with age (O'Connor and Parslow, 2009).

The CIDI screening questions for internalising disorders are relatively complex. Responding to them involves engaging several cognitive processes: attending to multiple components of the questions (e.g., "In the past 12 months, have you had two weeks or longer when nearly every day you felt sad, empty, or depressed for most of the day?"), processing several timeframes in working memory (e.g., "have you ever in your life had a period lasting several days or longer…") and retrieving autobiographical events from long-term memory. Decline in cognitive performance in old age particularly affects processing speed, working memory and episodic memory (Kennedy and Raz, 2009; Salthouse, 1996; Verhaeghen et al., 1993). Thus it is possible that older people are more sensitive to lengthy and complex assessment items and accordingly respond differently than younger people (O'Connor and Parslow, 2009; O'Connor and Parslow, 2010b). If this is the case then the screening items may be biased toward screening older adults out of the full diagnostic instrument, which could account, at least in part, for findings of lower prevalence rates of internalising disorders in old age.

The current study examined the extent to which methodological characteristics account for lower prevalence estimates of internalising disorders (i.e., unipolar mood and anxiety disorders) in older adults relative to middle-aged and younger adults. Specifically, are the screening questions of internalising disorders, enumerated in fully-structured epidemiological surveys, vulnerable to age-related bias? We examined this question using measurement invariance techniques.

Measurement invariance investigates whether items function equivalently amongst members of different groups (e.g., age, sex, diagnostic group) while accounting for the level of the underlying latent variable of interest. Establishing measurement invariance is thus a prerequisite for conducting meaningful group comparisons; lack of measurement invariance is indicative of systematic differences in response patterns between groups and precludes meaningful group comparisons.

In the context of the current study, measurement non-invariance would indicate that older adults who have the same level of internalising pathology as middle-aged and younger adults have different response patterns to the internalising screening items. In effect, this would suggest that related context-specific factors, such as question item complexity, may be artefactually contributing to the lesser rates of internalising pathology in older adults relative to younger adults. By contrast, if evidence of measurement invariance is found in the current study, this would suggest that older, middle-aged, and younger adults have equivalent response patterns, that group comparisons are meaningful, and, accordingly, that the reported decrease in prevalence of internalising disorders in old age is not attributable to bias in the screening items. As an alternative explanation, measurement error of a different kind (e.g., bias in the content of standardised diagnostic criteria) may account for the lower prevalence estimates or these estimates may in fact be an accurate reflection of mental health in older people. We predicted that measurement invariance would not be identified, based on previous

research suggesting older adults may be more sensitive to the cognitive demands of complicated screening questions than younger and middle-aged adults (O'Connor and Parslow, 2009; O'Connor and Parslow, 2010a).

Measurement invariance is typically tested by specifying and estimating a series of increasingly restricted multi-group confirmatory factor models and examining the change in fit between each model. The advantage of using this technique over simpler factor analyses is that it allows for direct comparisons of response patterns between the age groups: significant differences between models reflect systematic differences in the responses groups, or measurement non-invariance (Muthén and Asparouhov, 2002; van de Schoot et al., 2012). This study draws on data from three surveys: the 1997 and 2007 Australian National Survey of Mental Health and Wellbeing (NSMHWB), and the 2001-2002 National Epidemiologic Survey on Alcohol and Related Conditions (NESARC). Each of these surveys is based on a different diagnostic interview (see Method for further details), which provides a unique opportunity for investigating potential age-related bias and establishing that the results are robust.

2. Method

2.1. Sample

Data were drawn from three epidemiological surveys in the United States and Australia: the 1997 and 2007 Australian National Surveys of Mental Health and Wellbeing (NSMHWB), and the 2001-2002 National Epidemiologic Survey on Alcohol and Related Conditions (NESARC). All three surveys contained structured diagnostic interviews that were conducted face-to-face by trained lay-interviewers. The 1997 NSMHWB surveyed 10,641 community residents (78% response rate) using a stratified, multi-stage area probability sample of private households across urban and rural Australia. One resident aged 18 years or older was randomly selected from each household to participate in the interview.

The sample excluded institutions (e.g., aged care facilities and jails) as well as very remote or difficult to access regions of the country. The 2007 NSMHWB followed the same stratified sampling procedure as the 1997 survey and had 8,841 respondents (60% response rate). In the 2007 survey, the age range for interview participants was 16-85 years. As the youngest and oldest ages tend to be underrepresented using this procedure, the 2007 survey oversampled the youngest (16-24 years) and oldest (65-85 years) age groups to ensure reliable estimates. The 2001-2002 NESARC surveyed 43,093 non-institutional community residents (response rate 81%) in the US from a multi-stage stratified sample. Participants were aged 18 years and older, with young adults (18-24 years) oversampled to ensure sufficient representation of the youngest age group. The sampling frame was derived from US census data and included group-quarters sampling. Further information on the survey procedures and analyses of participant characteristics are described in previous studies (for further details on the 1997 and 2007 NSMHWB and NESARC, respectively, see Andrews et al., 2001; Slade et al., 2009; Grant et al., 2006). Data in each survey were weighted according to demographic distributions in the general population based on national census data. For this study, participants were categorised into three age groups: older adults (65 years and older), middle-aged (35-64 years) and younger adults (16-34 years). The number of participants across age groups and surveys are shown in Table 2.

2.2. Measures

The 1997 NSMHWB contained the Composite International Diagnostic Interview version 2.1 (CIDI 2.1), the 2007 NSHMWB was based on the CIDI World Mental Health version (WMH-CIDI), and the 2001-2002 NESARC used the Alcohol-Use Disorder and Associated Disabilities Interview Schedule-DSM-IV Version (AUDADIS-IV). The CIDI and AUDADIS-IV are both structured interviews designed to assess mental disorders according to ICD-10 and DSM-IV criteria (Andrews and Peters, 1998; Grant et al., 2003). All three

instruments are widely used in epidemiology studies and have strong psychometric properties (Grant et al., 2003; Jordanova et al., 2004; Kessler et al., 2004). The screening items in the 2007 NSMHWB and the NESARC queried lifetime occurrence of symptoms, while the 1997 NSMHWB assessed past-year symptomatology. The screening items analysed in this study (see Table 1) assessed major depression, generalised anxiety disorder, bipolar disorder, panic disorder, social phobia, and agoraphobia. In total, eight screening items were extracted from the WMH-CIDI (2007 NSMHWB), seven items from the CIDI 2.1 (1997 NSMHWB), and nine from the AUDADIS-IV (2001-2002 NESARC). Data were coded as dichotomous responses.

2.3. Statistical analyses

To investigate measurement invariance we estimated a series of increasingly restrictive multi-group confirmatory factor models based on guidelines in the extant literature (Muthén and Muthén, 2010). The advantage of multi-group confirmatory factor analysis is that it examines the factor models across different groups, allowing us to test for systematic differences in response patterns between age groups in each survey. First, we investigated the factor structure of the internalising screening items by applying exploratory and confirmatory factor analyses (EFA and CFA) across each age group and survey separately. Referred to as configural invariance or equal form, this step evaluates whether the basic factor structure is the same across groups in terms of the number of factors and whether the pattern of salient and non-salient factor loadings is identical. Configural invariance is the minimum condition necessary to enable group comparisons as it suggests that the constructs are conceptualised in a similar manner across groups (Carragher, 2008). In the configural model, factor structure was equal across age groups but loadings, thresholds and residual variances were allowed to vary.

Second, we estimated a scalar invariance model whereby factor loadings and thresholds were constrained to equality across age groups but residual variances were allowed to vary. Third, we tested a partial scalar model, in which loadings and thresholds of some items were allowed to vary between age groups based on significant and conceptually meaningful modification fit indices of the scalar model. Further details on this methodology are provided elsewhere (Muthén and Asparouhov, 2002; van de Schoot et al., 2012). All analyses were conducted in Mplus version 7 (Muthén and Muthén, 2010).

EFA models were conducted using the weighted least squares mean and variance adjusted estimator (WLSMV; Muthen and Muthen, 2010) and oblique rotation. Eigenvalues, scree plots, and model fit indices (described below) were used to assess model fit. CFA models were conducted using WLSMV estimation. Model fit was assessed using chi-square (χ^2) as well as other statistical fit indices which are less sensitive to large sample sizes (Browne et al., 2002). This included: the root mean square error of approximation (RMSEA) and its confidence intervals (Steiger, 1990), the Tucker-Lewis Index (TLI; Tucker and Lewis, 1973), and the Comparative Fit Index (CFI; Bentler, 1990). TLI and CFI values ≥0.90 indicate acceptable fit and values ≥0.95 imply very good fit (Hu and Bentler, 1998; Vandenberg and Lance, 2000). RMSEA values less than 0.05 indicate close model fit, values up to 0.08 suggest a reasonable error of approximation in the population, and values exceeding 0.10 indicate poor fit (Browne and Cudeck, 1993). McCallum et al. (1996) suggest that RMSEA values in the range of 0.08-1.0 indicate mediocre fit. Factor loadings (>=.0.40), correlations, and modification fit indices were also evaluated to identify the best fitting model.

3. Results

3.1. 2007 Australian National Survey of Mental Health and Wellbeing

EFA results for the 2007 NSMHWB suggested a one-factor solution provided the best

fit to the data for the WMH-CIDI internalising screening items. One component indexed an eigenvalue over one and the scree plot displayed a sharp decrease in eigenvalues between the first and second factor solutions. The factor loadings for all items were salient (0.756 – 0.953) and statistically significant (p<0.05). The one-factor solution provided very good model fit (CFI=0.971, TLI=0.962, RMSEA=0.058). Inspection of the two-factor solution revealed that only two items (representing social phobia and agoraphobia) loaded on the second factor. Ideally, several items should load on each factor in order to instil confidence that the latent construct under investigation is adequately measured and that the factor model is stable. The second factor solution also showed multiple item cross-loadings. Accordingly, the one-factor model was selected as the best-fitting model and separate one-factor CFA models were specified for each age group. These models demonstrated good fit to the data (see Table 2).

The goodness-of-fit results for the measurement invariance models are presented in Table 3. The χ^2 difference test indicated a significant difference in fit between the configural and scalar models, and no significant difference between the scalar and partial scalar models. As mentioned earlier however, the χ^2 statistic is highly sensitive to large sample sizes, so the CFI difference test and RMSEA 90% confidence intervals were also used to determine change in model fit. Changes in CFI values between successive models were trivial: ΔCFI = -0.001 from the configural to the scalar model, ΔCFI = -0.002 from the scalar to partial scalar model. The RMSEA 90% confidence intervals overlapped between the models. The negligible differences between the configural and scalar models indicated no systematic differences in how the different age groups responded to the CIDI internalising screening items. Further, the non-significant differences between the scalar and partial scalar models suggested that even when some of the constraints were allowed to vary, model fit was not substantially affected. In summary, these results collectively indicate that the CIDI screening items are invariant across older, middle-aged and young adults in the 2007 NSMHWB.

3.2. 1997 Australian National Survey of Mental Health and Wellbeing

EFA results for the 1997 NSMHWB suggested that a one-factor solution provided the best fit to the data for the CIDI 2.1 internalising screening items. One component indexed an eigenvalue over one and the scree plot displayed a sharp decrease in eigenvalues between the first and second factor solutions. The factor loadings for all items were salient (0.411 – 0.865) and statistically significant (p<0.05). The one-factor solution provided very good model fit (CFI=0.968, TLI=0.952, RMSEA=0.046). Inspection of the two-factor solution revealed that items representing anxiety, lost interest, and depressed mood loaded on a second factor. There were multiple cross-loadings in the two-factor model. Separate one-factor CFA models were specified for each age group and demonstrated an acceptable fit to the data (see Table 2).

The model fit statistics for the measurement invariance testing are shown in Table 3. No change in CFI values between successive models was observed and the RMSEA 90% confidence intervals overlapped between the models. In summary, these results provide support for measurement invariance across age groups in the 1997 NSMHWB for the CIDI internalising screening items.

3.3. 2001-2002 National Epidemiologic Survey on Alcohol and Related Conditions

EFA results for the 2001-2002 NESARC suggested a two-factor solution provided the best fit to the data for the AUDADIS-IV internalising screening items. Two components indexed eigenvalues over one and inspection of the scree plot supported a two-factor solution. The factor loadings for all items were salient (0.523 – 0.971) and statistically significant (p<0.05). The two-factor solution provided very good model fit (CFI=0.977, TLI=0.961, RMSEA=0.066) and was a substantial improvement compared to the one-factor solution (CFI=0.937, TLI=0.919, RMSEA=0.095). The two factors appeared to represent the distress and fear subfactors of internalising psychopathology reported elsewhere in the literature (e.g.,

Watson, 2005). The screening items for panic attacks and social phobia loaded onto one factor, while the items for depression, general anxiety and bipolar loaded onto the second factor. Separate two-factor CFA models were specified for each age group and demonstrated a good fit to the data (see Table 2).

The model fit statistics for the measurement invariance testing are shown in Table 3. No change in CFI was observed between successive models and the RMSEA 90% confidence intervals overlapped. In summary, these findings suggest that the AUDADIS-IV internalising screening items were invariant across age groups in the 2001-2002 NESARC.

4. Discussion

The purpose of the current study was to investigate age-related bias in the screening items of three widely used diagnostic instruments, the CIDI 2.1, WMH-CIDI and AUDADIS-IV. We predicted that evidence of measurement invariance between age groups would not be found, suggesting that systematic differences exist in the endorsement of internalising screening items between older, middle-aged, and young adults, which would reflect prevalence reports in the wider literature. Contrary to expectations, we found that the internalising screening items were invariant across age groups in each of the three national epidemiological surveys. Screening items are often included in large epidemiological surveys to reduce respondent burden and financial cost; the current findings lend support to the validity of using screening items for internalising disorders in the CIDI, WMH-CIDI, and AUDADIS-IV as they are free from age-related measurement bias.

The current results suggest that significantly lower endorsement of CIDI screening items by older adults (O'Connor and Parslow, 2010a) is not due to age-related bias in the items. This finding is in contrast to previous research that indicates older adults may respond differently than younger adults to cognitively complex items (O'Connor and Parslow, 2009; O'Connor and Parslow, 2010b), as cognitive decline in old age affects several processes

involved in responding to complicated questions about past autobiographical events (Kennedy and Raz, 2009; Salthouse, 1996; Verhaeghen et al., 1993). It is possible that O'Connor and Parslow's (2009) findings of increased disagreements between CIDI screening items and K-10 questions with age are a result of processes other than complexity of the CIDI items. For example the K-10 and the CIDI may measure different constructs that may change differently over the lifespan, such as psychological distress in the K-10 (Kessler et al., 2002) and pathology in the CIDI (Andrews and Peters, 1998).

A possible explanation for the current findings is a selection bias in the present samples, as a certain degree of cognitive functioning was required to respond to questions and complete the survey. This is necessary to the survey design, because administering interviews to people who exhibit pathological cognitive decline may lead to invalid responses. Indeed in the Australian surveys respondents aged 65+ years were screened for cognitive functioning using the Mini-Mental State Examination (Folstein et al., 1975), and respondents who scored 18 or less out of 30 did not participate in the full interview. However, the non-invariance finding was replicated in the NESARC, which did not have a screening component for cognitive functioning, demonstrating that the present results are not solely attributable to the exclusion of older people with lower cognitive capacity. Alternatively, it is possible that the degree of complexity in the CIDI screening items affected older and younger respondents equally. Both these explanations would suggest that the nature and extent of complexity in the CIDI screening items is not great enough to be susceptible to the minor cognitive decline that typically occurs in old age. To examine this further it would be beneficial for future research to use measures of impairment in specific cognitive domains in assessing changes in response, and to systematically investigate the processes that underlie older adults' responses to questions in diagnostic instruments using techniques such as cognitive interviewing.

The present findings also suggest that any differences in how internalising psychopathology is exhibited in old age are not evident in responses to the screening items. By design, the screening items reflect symptoms that most accurately represent the disorder and that have good concordance with full diagnoses (Kessler et al., 2012). Atypical presentations are therefore more likely to be better captured by full assessment rather than by screening items. For example, older adults often do not exhibit a typical pattern of depression symptoms; instead they may display fewer sadness-related symptoms while somatic symptoms, such as fatigue and sleep disturbances, are more pronounced (Gallo and Rabins, 1999; Gonçalves et al., 2009; Hegeman et al., 2012). As the CIDI and AUDADIS-IV screening items across the mood and anxiety disorders do not represent these somatic symptoms, it is possible that older adults may be disproportionately screened out of the full assessment if they have different core symptoms to those represented in the screening questions. This would account, at least in part, for previous findings of a decline in prevalence of internalising disorders in older adults as well as the current results since it would mean differences in symptom profiles are not captured by the screening items. Further research examining age-related differences in response patterns to diagnostic questions across disorders in the full instruments would be beneficial to elucidate these findings. Unfortunately these analyses are often precluded in community samples, as large epidemiological surveys tend to include skip questions to reduce respondent burden and financial costs.

Another observation to emerge from the study is the finding that a one-factor model provided the best fit to the CIDI screening items, while a two-factor model provided the best fit to the AUDADIS-IV data. Numerous studies have examined the underlying structure of psychopathology and yielded robust support for the internalising dimension. Notably, some of these studies have found evidence for an internalising dimension (e.g. Krueger et al., 1998)

whereas others have demonstrated that the internalising dimension may be bifurcated into anxious-misery and fear subdimensions (Eaton et al., 2011; Krueger, 1999; Slade and Watson, 2006; Vollebergh et al., 2001). The two factors found for the AUDADIS-IV screening items represented anxiety-misery (the symptoms for major depression and general anxiety; bipolar symptoms also loaded on this factor) and fear (symptoms for panic disorder and social phobia). It is possible that the AUDADIS-IV screening items may be more sensitive to the two subfactors than the CIDI, and may provide more discrimination on the latent internalising dimension for determining whether fear or distress disorders are present from the screening items. It is important to note that this would not affect case ascertainment for either instrument, as diagnostic information for the purpose of estimating prevalence rates is based on the full diagnostic assessment.

A limitation of this study relates to the number and conceptualisation of the screening items, as the wording of the items and the number of different symptoms represented differs across the three diagnostic instruments. These differences make it difficult to directly compare models between survey administrations and may explain the different factor structures found for the CIDI and AUDADIS-IV. Across surveys, however, all items have similar complexities that may be problematic for people with cognitive decline as they all require recall and manipulation of multiple types of information. Further, examining three epidemiological surveys is a notable strength of this study, and finding evidence for measurement invariance in all three instruments indicates that the present results are robust.

Another limitation is that the national surveys analysed in this study excluded residents of aged-care facilities. This exclusion criterion was necessary as there is a high prevalence of dementia among aged-care residents, and also the facilities are not randomly distributed geographically (O'Connor, 2006). The prevalence of internalising disorders is estimated to be relatively high in Australian nursing homes; therefore excluding residents

from the sample may affect age-related differences in prevalence estimates in a community sample, particularly among the 'old-old' adults (Jorm, 2000; O'Connor, 2006). A further limitation is that the older adult age group of 65+ years is likely heterogeneous in respect to long-term memory retrieval and cognitive decline. There may be more fine-grained age differences in response patterns within the older adult age group; however these could not be examined using the current method as the 2007 NSMHWB only sampled up to age 85 years and numbers are relatively low in the older ages. Moreover, age-related declines in memory processes and processing speed are apparent from at least early old age (e.g., Hedden & Gabrieli, 2004; Nilsson, 2003; Öztekin, Güngör & Badre, 2012; Verhaeghen et al., 1993), so if there are global effects of slowed cognition on the ability to answer screening questions we would still expect to see differences across the older adult age group compared to younger adults.

The present study represents an important first step in systematically evaluating whether reported prevalence differences of internalising disorders amongst older adults is valid or the result of methodological bias. We found that low prevalence estimates of internalising disorders in older adults are unlikely attributable to age-group differences in endorsement patterns of the screening items in the CIDI and AUDADIS-IV. That being said, the findings do not preclude the possibility that age-related bias may be present in the remaining items in the respective diagnostic modules for each of the internalising disorders. For example, older people may be more likely than younger people to attribute depressive symptoms to physiological effects, which may in turn be used as a strategy for simplifying recall processes in diagnostic interviews (Knäuper and Wittchen, 1994). Further testing is required to determine if there are differences in how older adults present internalising psychopathology. That is, are they more likely to experience and report somatic symptoms,

resulting in systematic differences in endorsement of diagnostic items and therefore differences in prevalence estimates?

In closing, this study found evidence of measurement invariance in the screening items for internalising disorders in three widely used epidemiology assessment measures, the WMH-CIDI, CIDI 2.1, and AUDADIS-IV. This offers support to the validity of these instruments, specifically for the use of the screening items, as it suggests that no substantial differences are present in endorsement patterns amongst older, middle-aged, and young adults. The invariance of the screening items suggests that the lower overall endorsement of these items by older people (O'Connor and Parslow, 2010a) is not attributable to differential response patterns across items, and therefore older adults are not differentially screened out of the full diagnostic instrument. Future analyses should be targeted towards evaluating the presence of age-related bias in responses to items in the entire mood and anxiety disorders modules in order to conclusively determine whether rates of mental disorders truly decline in old age or are a function of methodological bias.

References

Andrews, G., Henderson, S., Hall, W., 2001. Prevalence, comorbidity, disability and service utilisation: Overview of the Australian National Mental Health Survey. The British Journal of Psychiatry 178, 145-153. doi: 10.1192/bjp.178.2.145

Andrews, G., Peters, L., 1998. The psychometric properties of the Composite International Diagnostic Interview. Social Psychiatry and Psychiatric Epidemiology 33, 80-88. doi: 10.1007/s001270050026

Beekman, A.T.F., Bremmer, M.A., Deeg, D.J.H., Van Balkom, A.J.L.M., Smit, J.H., De Beurs, E., Van Dyck, R., Van Tilburg, W., 1998. Anxiety disorders in later life: A report from the Longitudinal Aging Study Amsterdam. International Journal of Geriatric Psychiatry 13, 717-726. doi: 10.1002/(sici)1099-1166(1998100)13:10<717::aid-gps857>3.0.co;2-m

Bentler, P.M., 1990. Comparative fit indexes in structural models. Psychological Bulletin 107, 238-246.

Blazer, D.G., 2003. Depression in late life: Review and commentary. The Journals of Gerontology Series A: Biological Sciences and Medical Sciences 58, M249-M265. doi: 10.1093/gerona/58.3.M249

Blazer, D.G., Hybels, C.F., 2005. Origins of depression in later life. Psychological Medicine 35, 1241-1252. doi: 10.1017/S0033291705004411

Browne, M.W., Cudeck, R. (1993). Alternative ways of assessing model fit. In Bollen, K.A. and Long, J.S. (Eds.), Testing Structural Equation Models. Beverly Hills, CA, Sage.

Browne, M.W., MacCallum, R.C., Kim, C.T., Anderson, B.L., Glaser, R., 2002. When fit indices and the residuals are incompatible. Psychological Methods 7, 403-421.

Carragher, N. (2008). Elucidating heterogeneity in depression symptom indicators using data from a large, nationally representative U.S. survey. PhD thesis, University of Ulster,

Northern Ireland.

Cheung, G.W., Rensvold, R.B., 1999. Testing factorial invariance across groups: A reconceptualization and proposed new method. Journal of Management 25, 1-27. doi: 10.1177/014920639902500101

Doraiswamy, P.M., Khan, Z.M., Donahue, R.M.J., Richard, N.E., 2002. The spectrum of quality-of-life impairments in recurrent geriatric depression. The Journals of Gerontology Series A: Biological Sciences and Medical Sciences 57, M134-M137. doi: 10.1093/gerona/57.2.M134

Eaton, N.R., Keyes, K.M., Balsis, S., Skodol, A.E., Markon, K.E., Grant, B.F., Hasin, D.S., 2012. An invariant dimensional liability model of gender differences in mental disorder prevalence. Journal of Abnormal Psychology 121.

Ernst, G., Angst, J., 1995. Depression in old age. Is there a real decrease in prevalence? A review. European Archives of Psychiatry and Clinical Neuroscience 245, 272-287.

Folstein, M.F., Folstein, S.E., McHugh, P.R., 1975. 'Mini-Mental State': A practical method for grading the cognitive state of patients for the clinician. Journal of Psychiatric Research 12, 189-198.

Gallo, J.J., Anthony, J.C., Muthen, B.O., 1994. Age differences in the symptoms of depression: A latent trait analysis. Journal of Gerontology 49, P251-P264. doi: 10.1093/geronj/49.6.P251

Gallo, J.J., Rabins, P.V., 1999. Depression without sadness: Alternative presentations of depression in late life. American Family Physician 60, 820-826.

Gonçalves, D.C., Albuquerque, P.B., Byrne, G.J., Pachana, N.A., 2009. Assessment of depression in aging contexts: General considerations when working with older adults. Professional Psychology: Research and Practice 40, 609-616. doi: 10.1037/a0017305

Grant, B.F., Dawson, D.A., Stinson, F.S., Chou, P.S., Kay, W., Pickering, R., 2003. The

Alcohol Use Disorder and Associated Disabilities Interview Schedule-IV (AUDADIS-IV): Reliability of alcohol consumption, tobacco use, family history of depression and psychiatric diagnostic modules in a general population sample. Drug and Alcohol Dependence 71, 7-16. doi: 10.1016/S0376-8716(03)00070-X

Grant, B.F., Stinson, F.S., Dawson, D.A., Chou, S.P., June Ruan, W., Pickering, R.P., 2006. Co-occurrence of 12-month alcohol and drug use disorders and personality disorders in the United States. Alcohol Research and Health 29, 121-130.

Hedden, T., Gabrieli, J.D.E., 2004. Insights into the ageing mind: A view from cognitive neuroscience. Nature Reviews Neuroscience 5, 87-96. doi: 10.1038/nrn1323

Hegeman, J.M., Wardenaar, K.J., Comijs, H.C., de Waal, M.W.M., Kok, R.M., van der Mast, R.C., 2012. The subscale structure of the Inventory of Depressive Symptomatology Self Report (IDS-SR) in older persons. Journal of Psychiatric Research 46, 1383-1388. doi: 10.1016/j.jpsychires.2012.07.008

Henderson, A.S., Jorm, A.F., Korten, A.E., Jacomb, P., Christensen, H., Rodgers, B., 1998. Symptoms of depression and anxiety during adult life: evidence for a decline in prevalence with age. Psychological Medicine 28, 1321-1328.

Hu, L.T., Bentler, P.M., 1998. Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. Psychological Methods 3, 424-453.

Jordanova, V., Wickramesinghe, C., Gerada, C., Prince, M., 2004. Validation of two survey diagnostic interviews among primary care attendees: A comparison of CIS-R and CIDI with SCAN ICD-10 diagnostic categories. Psychological Medicine 34, 1013-1024. doi: 10.1017/S0033291703001727

Jorm, A.F., 2000. Does old age reduce the risk of anxiety and depression? A review of epidemiological studies across the adult life span. Psychological Medicine 30, 11-22.

Kennedy, K.M., Raz, N., 2009. Aging white matter and cognition: Differential effects of

regional variations in diffusion properties on memory, executive functions, and speed. Neuropsychologia 47, 916-927. doi: 10.1016/j.neuropsychologia.2009.01.001

Kessler, R.C., Abelson, J., Demler, O., Escobar, J.I., Gibbon, M., Guyer, M.E., Howes, M.J., Jin, R., Vega, W.A., Walters, E.E., Wang, P., Zaslavsky, A., Zheng, H., 2004. Clinical calibration of DSM-IV diagnoses in the World Mental Health (WMH) version of the World Health Organization (WHO) Composite International Diagnostic Interview (WMH-CIDI). International Journal of Methods in Psychiatric Research 13, 122-139. doi: 10.1002/mpr.169

Kessler, R.C., Birnbaum, H., Bromet, E., Hwang, I., Sampson, N., Shahly, V., 2010a. Age differences in major depression: Results from the National Comorbidity Survey Replication (NCS-R). Psychological Medicine 40, 225-237. doi: 10.1017/S0033291709990213

Kessler, R.C., Birnbaum, H.G., Shahly, V., Bromet, E., Hwang, I., McLaughlin, K.A., Sampson, N., Andrade, L.H., de Girolamo, G., Demyttenaere, K., Haro, J.M., Karam, A.N., Kostyuchenko, S., Kovess, V., Lara, C., Levinson, D., Matschinger, H., Nakane, Y., Browne, M.O., Ormel, J., Posada-Villa, J., Sagar, R., Stein, D.J., 2010b. Age differences in the prevalence and co-morbidity of DSM-IV major depressive episodes: Results from the WHO World Mental Health Survey Initiative. Depression and Anxiety 27, 351-364. doi: 10.1002/da.20634

Kessler, R.C., Calabrese, J.R., Farley, P.A., Gruber, M.J., Jewell, M.A., Katon, W., Keck, P.E.J., Nierenberg, A.A., Sampson, N.A., Shear, M.K., Shillington, A.C., Stein, M.B., Thase, M.E., Wittchen, H.-U., 2012. Composite International Diagnostic Interview screening scales for DSM-IV anxiety and mood disorders. Psychological Medicine 18, 1-13. doi: 10.1017/S0033291712002334

Knäuper, B., Wittchen, H.-U., 1994. Diagnosing major depression in the elderly: Evidence

for response bias in standardized diagnostic interviews? Journal of Psychiatric Research 28, 147-164. doi: 10.1016/0022-3956(94)90026-4

Krueger, R.F., 1999. The structure of common mental disorders. Archives of General Psychiatry 56, 921-926.

Krueger, R.F., Caspi, A., Moffitt, T.E., Silva, P.A., 1998. The structure and stability of common mental disorders (DSM-III-R): A longitudinal-epidemiological study. Journal of Abnormal Psychology 107, 216-227.

Krueger, R.F., Markon, K.E., 2006. Reinterpreting comorbidity: A model-based approach to understanding and classifying psychopathology. Annual Review of Clinical Psychology 2, 111-133. doi: 10.1146/annurev.clinpsy.2.022305.095213

Lenze, E.J., Rogers, J.C., Martire, L.M., Mulsant, B.H., Rollman, B.L., Dew, M.A., Schulz, R., Reynolds, C.F., 2001. The association of late-life depression and anxiety with physical disability: A review of the literature and prospectus for future research. American Journal of Geriatric Psychiatry 9, 113-135.

McEvoy, P.M., Grove, R., Slade, T., 2011. Epidemiology of anxiety disorders in the Australian general population: Findings of the 2007 Australian National Survey of Mental Health and Wellbeing. Australian and New Zealand Journal of Psychiatry 45, 957-967. doi: 10.3109/00048674.2011.624083

Muthén, B.O., Asparouhov, T. (2002). Latent variable analysis with categorical outcomes: Multiple-group and growth modeling in Mplus. Mplus web notes, no 4 version 5. Retrieved from https://www.statmodel.com/download/webnotes/CatMGLong.pdf

Muthén, L.K., Muthén, B.O. (2010). Mplus Users' Guide. Los Angeles, CA, Muthén & Muthén.

O'Connor, D.W., 2006. Do older Australians truly have low rates of anxiety and depression? A critique of the 1997 National Survey of Mental Health and Wellbeing. Australian

and New Zealand Journal of Psychiatry 40, 623-631. doi: 10.1080/j.1440-1614.2006.01861.x

O'Connor, D.W., Parslow, R.A., 2009. Different responses to K-10 and CIDI suggest that complex structured psychiatric interviews underestimate rates of mental disorder in old people. Psychological Medicine 39, 1527-1531. doi: 10.1017/S0033291708004728

O'Connor, D.W., Parslow, R.A., 2010a. Differences in older people's responses to CIDI's depression screening and diagnostic questions may point to age-related bias. J. Affect. Disord. 125, 361-364. doi: 10.1016/j.jad.2010.01.008

O'Connor, D.W., Parslow, R.A., 2010b. Mental health scales and psychiatric diagnoses: Responses to GHQ-12, K-10 and CIDI across the lifespan. J. Affect. Disord. 121, 263-267. doi: 10.1016/j.jad.2009.06.038

Öztekin, I., Güngör, N.Z., Badre, D., 2012. Impact of aging on the dynamics of memory retrieval: A time-course analysis. Journal of Memory and Language 67, 285-294. doi: 10.1016/j.jml.2012.05.003

Riedel-Heller, S.G., Busse, A., Angermeyer, M.C., 2006. The state of mental health in old-age across the 'old' Eurpean Union - a systematic review. Acta Psychiatrica Scandinavica 113, 388-401. doi: 10.1111/j.1600-0447.2005.00632.x

Salthouse, T.A., 1996. The processing-speed theory of adult age differences in cognition. Psychological Review 103, 403-428. doi: 10.1037/0033-295x.103.3.403

Schulz, R., Drayer, R.A., Rollman, B.L., 2002. Depression as a risk factor for non-suicide mortality in the elderly. Biological Psychiatry 52, 205-225. doi: 10.1016/S0006-3223(02)01423-3

Slade, T., Johnston, A., Oakley Browne, M.A., Andrews, G., Whiteford, H., 2009. 2007 National Survey of Mental Health and Wellbeing: Methods and key findings.

Australian and New Zealand Journal of Psychiatry 43, 594-605. doi: 10.1080/00048670902970882

Slade, T., Watson, D., 2006. The structure of common DSM-IV and ICD-10 mental disorders in the Australian general population. Psychological Medicine 36, 1593-1600. doi: 10.1017/S0033291706008452

Steiger, J.H., 1990. Structural model evaluation and modification: An interval estimation approach. Multivariate Behavioral Research 25, 173-180. doi: 10.1207/s15327906mbr2502_4

Trollor, J.N., Anderson, T.M., Sachdev, P.S., Brodaty, H., Andrews, G., 2007. Age shall not weary them: Mental health in the middle-aged and the elderly. Australian and New Zealand Journal of Psychiatry 41, 581-589. doi: 10.1080/00048670701392817

United Nations Department of Economic and Social Affairs, P.D. (2002). Word Population Ageing 1950-2050. New York, UN DESA.

van de Schoot, R., Lugtig, P., Hox, J., 2012. Developmetrics: A checklist for testing measurement invariance. European Journal of Developmental Psychology, 1-7. doi: 10.1080/17405629.2012.686740

Vandenberg, R.J., Lance, C.E., 2000. A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. Organizational Research Methods 3, 4-70.

Verhaeghen, P., Marcoen, A., Goossens, L., 1993. Facts and fiction about memory aging: A quantitative integration of research findings. Journal of Gerontology 48, P157-P171. doi: 10.1093/geronj/48.4.P157

Volkert, J., Schulz, H., Härtera, M., Wlodarczyk, O., Andreas, S., 2013. The prevalence of mental disorders in older people in Western countries - a meta-analysis. Ageing Research Reviews 12, 339-353. doi: 10.1016/j.arr.2012.09.004

Vollebergh, W.A.M., Iedema, J., Bijl, R.V., de Graaf, R., Smit, F., Ormel, J., 2001. The structure and stability of common mental disorders: The NEMESIS study. Archives of General Psychiatry 58, 597-603. doi: 10.1001/archpsyc.58.6.597

Wetherell, J.L., Thorp, S.R., Patterson, T.L., Golshan, S., Jeste, D.V., Gatz, M., 2004. Quality of life in geriatric generalized anxiety disorder: A preliminary investigation. Journal of Psychiatric Research 38, 305-312. doi: 10.1016/j.jpsychires.2003.09.003

Wright, A.G.C., Krueger, R.F., Hobbs, M.J., Markon, K.E., Eaton, N.R., Slade, T., 2013. The structure of psychopathology: Toward an expanded quantitative empirical model. Journal of Abnormal Psychology 122, 281-294. doi: 10.1037/a0030133

Table 1
Screening items for mood and anxiety disorders in the WMH-CIDI (used in the 2007 NSMHWB), CIDI 2.1 (1997 NSMHWB) and AUDADIS-IV (2001-2002 NESARC).

Internalising disorder	Screening item
WMH-CIDI:	
(2007 NSMHWB)	
depression	Have you ever in your life had a period lasting several days or longer when most of the day you felt sad, empty, or depressed?
depression	Have you ever had a period lasting several days or longer when most of the day you were very discouraged about how things were going in your life?
depression	Have you ever had a period lasting several days or longer when you lost interest in most things you usually enjoy like work, hobbies, and personal relationships?
general anxiety	Did you ever have a time in your life when you were a "worrier" – that is, when you worried a lot more about things than other people with the same problems as you?
panic	Have you ever in your life had an attack of fear or panic when all of a sudden you felt very frightened, anxious, or uneasy?
social phobia	Looking at page 32 in your booklet, was there ever a time in your life when you felt very afraid or really, really shy with people, like meeting new people, going to parties, going on a date, or using a public bathroom?
agoraphobia	Looking at the bottom of page 32 in your booklet, was there ever a time in your life when you felt afraid of either being in crowds, going to public places, travelling by yourself, or travelling away from home?
bipolar	Some people have periods lasting four days or longer when they feel much more excited and full of energy than usual. Their minds go too fast. They talk a lot. They are very restless or unable to sit still and they sometimes do things that are unusual for them, such as driving too fast or spending too much money. Have you ever had a period like this lasting several days or longer?
bipolar	Have you ever had a period lasting four days or longer when most of the time you were very irritable, grumpy, or in a bad mood?
CIDI 2.1:	
(1997 NSMHWB)	
depression	Now I want to ask you about periods of feeling sad, empty, or depressed. In the past 12 months, have you had two weeks or longer when nearly every day you felt sad, empty, or depressed for most of the day?
depression	In the past 12 months, have you had 2 weeks or longer when you lost interest in most things like work, hobbies, and other things you usually enjoyed?
dysthymia	Have you ever had two years or more in your life when you felt depressed or sad most days, even if you felt OK sometimes?
general anxiety	Now I want to ask you about longer periods of feeling worried, tense or anxious. In the past 12 months, have you had a period of a month or more when most of the time you felt worried, tense or anxious, about everyday problems such as work or family?
panic	Now I would like to ask you about attacks of fear that could happen anywhere. In your entire lifetime, have you ever had an attack when all

	of a sudden you felt frightened, anxious or very uneasy?
social phobia	Now I would like to ask you about situations in which you may have been anxious or afraid. Some people have a strong fear of doing things in front of others or of being the centre of attention. Look at the situations on Card D2. In the past 12 months, have you had an unusually strong fear or unusually strong avoidance of any of the situations on the list?
specific phobia	Now I would like to ask you about other situations in which you may have been afraid. Look at the situations on Card D3. In the past 12 months, have you had an unusually strong fear or unusually strong avoidance of any of the situations on the list?
AUDADIS-IV: *(2001-2002 NESARC)*	
depression	In your entire life, have you ever had a time when you felt sad, blue, depressed, or down most of the time for at least 2 weeks?
depression	In your entire life, have you ever had a time, lasting at least 2 weeks, when you didn't care about the things that you usually cared about, or when you didn't enjoy the things you usually enjoyed?
general anxiety	Have you ever had a time lasting at least 6 months when you felt tense, nervous, or worried most of the time?
panic	Have you ever had a panic attack, when all of a sudden you felt frightened, overwhelmed or nervous, almost as if you were in great danger, but really weren't?
panic	Did you ever think you were having a heart attack, but the doctor said it was just nerves or you were having a panic attack?
social phobia	Some people have such a strong fear of social situations, like doing things in front of other people or being the center of attention, that they become very frightened and nervous or they try to avoid them. Did you ever have such a strong fear or avoidance of any social situation
social phobia	Did you ever have a strong fear or avoidance of any social situation because you were afraid you would become speechless, have nothing to say or you might say something foolish?
bipolar	In your entire life, have you ever had a time lasting at least 1 week when you felt so extremely excited, elated or hyper that other people thought you weren't your normal self?
bipolar	In your entire life, have you ever had a time lasting at least 1 week when you were so irritable or easily annoyed that you would shout at people, throw or break things, or start fights or arguments?

Notes: WMH-CIDI = Composite International Diagnostic Interview, World Mental Health Version. CIDI 2.1 = Composite International Diagnostic Interview version 2.1. AUDADIS-IV = Alcohol Use Disorder and Associated Disabilities Interview Schedule–DSM-IV Version. NSMHWB = Australian National Survey of Mental Health and Wellbeing. NESARC = National Epidemiologic Survey on Alcohol and Related Conditions.

Table 2
Confirmatory factor analyses fit indices across age groups in the 2007 and 1997 Australian National Survey of Mental Health and Wellbeing (NSMHWB) and 2001-2002 National Epidemiologic Survey on Alcohol and Related Conditions (NESARC).

	N	χ^2	df	*p*	CFI	TLI	RMSEA
2007 NSMHWB							
Younger adults	2761	399	27	<0.001	0.965	0.954	0.071
Middle-aged adults	4175	321	27	<0.001	0.978	0.971	0.051
Older adults	1905	167	27	<0.001	0.968	0.957	0.052
Total sample	8827	816	27	<0.001	0.971	0.962	0.058
1997 NSMHWB							
Younger adults	3258	89.479	14	<0.001	0.975	0.962	0.041
Middle-aged adults	5591	244.829	14	<0.001	0.963	0.945	0.054
Older adults	1792	61.719	14	<0.001	0.942	0.913	0.044
Total sample	10641	334.671	14	<0.001	0.968	0.952	0.046
NESARC							
Younger adults	12958	1508	34	<0.001	0.975	0.967	0.058
Middle-aged adults	21930	2633	34	<0.001	0.973	0.964	0.059
Older adults	8205	661	34	<0.001	0.972	0.963	0.047
Total sample	43093	4868	34	<0.001	0.968	0.958	0.057

Notes: χ^2 = chi-square goodness-of-fit index, df = degrees of freedom, *p* = significance of chi-square value, CFI = comparative fit index, TLI = Tucker-Lewis index, RMSEA = root mean square error of approximation.
2007 and 1997 NSMHWB results relate to the one-factor models; NESARC results correspond to the two-factor model.

Table 3
Multi-group CFA measurement invariance model fit indices for the 2007 and 1997 Australian National Survey of Mental Health and Wellbeing (NSMHWB) and 2001-2002 National Epidemiologic Survey on Alcohol and Related Conditions (NESARC).

	#	CFI	ΔCFI	RMSEA	90% CI	$\Delta\chi^2$	df	*p*
2007 NSMHWB								
Configural invariance	54	0.971	-	0.059	0.055-0.062	-	-	-
Scalar invariance	40	0.972	-0.001	0.053	0.050-0.056	52.85	14	<0.001
Partial scalar invariance[a]	46	0.973	-0.002	0.054	0.051-0.057	9.583	8	0.296
1997 NSMHWB								
Configural invariance	42	0.967	-	0.048	0.044-0.052	-	-	-
Scalar invariance	32	0.967	0.000	0.043	0.039-0.047	50.960	10	<0.001
Partial scalar invariance[b]	36	0.967	0.000	0.044	0.040-0.049	25.307	6	<0.001
2001-2002 NESARC								
Configural invariance	63	0.975	-	0.056	0.054-0.057	-	-	-
Scalar invariance	51	0.975	0.000	0.053	0.052-0.054	195.592	12	<0.001
Partial scalar invariance[c]	57	0.975	0.000	0.054	0.053-0.056	16.852	6	0.01

[a] Partial invariance involved freeing loadings and thresholds for the items representing depressed mood, anxiety, and agoraphobia.
[b] Partial invariance involved freeing loadings and thresholds for items representing lost interest and depressed mood.
[c] Partial invariance model: free loadings and thresholds for items representing mania, irritability/anger, and panic attack.
Notes: # = number of free parameters, CFI = comparative fit index, ΔCFI = change in CFI value between successive models, RMSEA = root mean square error of approximation, 90% CI = 90% confidence intervals of RMSEA, $\Delta\chi^2$ = change in chi-square value between successive models, df = degrees of freedom, *p* = significance of the chi-square difference test.