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

Author/Contributor:
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

License:
https://creativecommons.org/licenses/by-nc-nd/3.0/au/
Link to license to see what you are allowed to do with this resource.

Downloaded from http://hdl.handle.net/1959.4/53133 in https://unswworks.unsw.edu.au on 2023-09-14
Investigating age-related differences in responses to screening items for internalising disorders in three national surveys

Heather Buchan
Matthew Sunderland
Natacha Carragher
Philip Batterham
Tim Slade

Manuscript in preparation for Journal of Affective Disorders

1NHMRC 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

2Centre 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 (χ²) 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 $\chi^2$ 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 $\chi^2$ 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: $\Delta$CFI = -0.001 from the configural to the scalar model, $\Delta$CFI = -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.,
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


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


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


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).

<table>
<thead>
<tr>
<th>Internalising disorder</th>
<th>Screening item</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>WMH-CIDI: (2007 NSMHWB)</strong></td>
<td></td>
</tr>
<tr>
<td>depression</td>
<td>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?</td>
</tr>
<tr>
<td>depression</td>
<td>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?</td>
</tr>
<tr>
<td>depression</td>
<td>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?</td>
</tr>
<tr>
<td>general anxiety</td>
<td>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?</td>
</tr>
<tr>
<td>panic</td>
<td>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?</td>
</tr>
<tr>
<td>social phobia</td>
<td>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?</td>
</tr>
<tr>
<td>agoraphobia</td>
<td>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?</td>
</tr>
<tr>
<td>bipolar</td>
<td>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?</td>
</tr>
<tr>
<td>CIDI 2.1: (1997 NSMHWB)</td>
<td></td>
</tr>
<tr>
<td>depression</td>
<td>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?</td>
</tr>
<tr>
<td>depression</td>
<td>In the past 12 months, have you had 2 weeks or longer when you lost interest in most things you usually enjoy?</td>
</tr>
<tr>
<td>dysthymia</td>
<td>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?</td>
</tr>
<tr>
<td>general anxiety</td>
<td>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?</td>
</tr>
</tbody>
</table>
| 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
<table>
<thead>
<tr>
<th>Condition</th>
<th>Question</th>
</tr>
</thead>
<tbody>
<tr>
<td>social phobia</td>
<td>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?</td>
</tr>
<tr>
<td>specific phobia</td>
<td>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?</td>
</tr>
<tr>
<td>depression</td>
<td>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?</td>
</tr>
<tr>
<td>general anxiety</td>
<td>Have you ever had a time lasting at least 6 months when you felt tense, nervous, or worried most of the time?</td>
</tr>
<tr>
<td>panic</td>
<td>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?</td>
</tr>
<tr>
<td>social phobia</td>
<td>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 because you were afraid you would become speechless, have nothing to say or you might say something foolish?</td>
</tr>
<tr>
<td>bipolar</td>
<td>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?</td>
</tr>
<tr>
<td>bipolar</td>
<td>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?</td>
</tr>
</tbody>
</table>

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
Confirmaory 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).

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

Notes: \( \chi^2 \) = 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).

<table>
<thead>
<tr>
<th></th>
<th>#</th>
<th>CFI</th>
<th>ΔCFI</th>
<th>RMSEA</th>
<th>90% CI</th>
<th>Δχ²</th>
<th>df</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>2007 NSMHWB</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Configural invariance</td>
<td>54</td>
<td>0.971</td>
<td>-</td>
<td>0.059</td>
<td>0.055-0.062</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Scalar invariance</td>
<td>40</td>
<td>0.972</td>
<td>-0.001</td>
<td>0.053</td>
<td>0.050-0.056</td>
<td>52.85</td>
<td>14</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Partial scalar invariance a</td>
<td>46</td>
<td>0.973</td>
<td>-0.002</td>
<td>0.054</td>
<td>0.051-0.057</td>
<td>9.583</td>
<td>8</td>
<td>0.296</td>
</tr>
<tr>
<td><strong>1997 NSMHWB</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Configural invariance</td>
<td>42</td>
<td>0.967</td>
<td>-</td>
<td>0.048</td>
<td>0.044-0.052</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Scalar invariance</td>
<td>32</td>
<td>0.967</td>
<td>0.000</td>
<td>0.043</td>
<td>0.039-0.047</td>
<td>50.960</td>
<td>10</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Partial scalar invariance b</td>
<td>36</td>
<td>0.967</td>
<td>0.000</td>
<td>0.044</td>
<td>0.040-0.049</td>
<td>25.307</td>
<td>6</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td><strong>2001-2002 NESARC</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Configural invariance</td>
<td>63</td>
<td>0.975</td>
<td>-</td>
<td>0.056</td>
<td>0.054-0.057</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Scalar invariance</td>
<td>51</td>
<td>0.975</td>
<td>0.000</td>
<td>0.053</td>
<td>0.052-0.054</td>
<td>195.592</td>
<td>12</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Partial scalar invariance c</td>
<td>57</td>
<td>0.975</td>
<td>0.000</td>
<td>0.054</td>
<td>0.053-0.056</td>
<td>16.852</td>
<td>6</td>
<td>0.01</td>
</tr>
</tbody>
</table>

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, Δχ² = change in chi-square value between successive models, df = degrees of freedom, p = significance of the chi-square difference test.