Comparing depression diagnostic symptoms across younger and older adults

Steve Balsis* and Jeffrey A. Cully

“Texas A & M University, College Station, USA; Baylor College of Medicine, Houston, USA

(Objective: Depression in later life has potential grave implications and contributes to heavy emotional, medical, and economic burdens. Therefore, it is not surprising that identifying depression and its symptoms in later life has remained a sustained concern for professionals who treat older patients. Despite this concern, the current diagnostic gold standard may not identify depression symptoms equally well in older and younger adults. The objective of this analysis is to determine whether older and younger adults with equivalent levels of latent depression are equally likely to endorse particular DSM diagnostic symptoms. Method: We analyzed DSM depression data using Item Response Theory (IRT)-based differential item functioning analyses. The data came from 1808 older adults (age 65–98 years) and 3734 younger adults (age 18–34 years) who participated in the National Epidemiological Survey on Alcoholism and Related Conditions. Results: The analyses confirmed our hypothesis. The DSM items identify depression differently in younger and older adults. Specifically, results showed that older adults were more likely to endorse somatic items and less likely to endorse cognitive and suicide items than their younger counterparts with equivalent levels of depression. Conclusion: These findings provide evidence that the DSM depression items work differently across age groups when controlling for latent depression. It is important to consider, however, that these findings are limited by the sampling methodology and the particular protocol implemented.

Keywords: age; aging; DSM; item response theory

Introduction

The accurate measurement of depression in later life has been a concern for gerontologists (e.g., Blazer, 2002). This concern may be because of the well-documented link between depression in later life and medical diseases such as cardiovascular disease and diabetes (Cosgrove, Sargeant, & Griffin, 2008; Whooley, 2006), neurological diseases such as Alzheimer’s disease and stroke (Arbelaez, Ariyo, Crum, Fried, & Ford, 2007; Ownby, Acevedo, Harwood, Barker, & Duara, 2008), lowered function (Ormel, Rijsdijk, Sullivan, van Sonderen, & Kempen, 2002), mortality (Ryan et al., 2008), and suicide (Crocker, Clare, & Evans, 2006). The implications of depression in later life can be grave, yet the DSM items (American Psychiatric Association, 1987, 1994, 2000) that help to define the disorder may mischaracterize it in older adults.

Many have long suspected that several DSM items originally written to identify depression in younger adults may not identify depression equally well in older adults (Jeste, Blazer, & First, 2007). Intuitively, the somatic and cognitive depression items in the DSM should work less well to characterize depression in later life. As an example, somatic symptoms such as sleep problems can occur because of medical complications (e.g., Rudy, Weiner, Lieber, Slaboda, & Boston, 2007), medication side-effects, and the natural sleep changes that accompany aging (e.g., Giganti, Ficca, & Salzarulo, 2008). One might suspect that an item that identifies sleep problems would be endorsed more frequently by older adults when compared to younger adults with an equivalent level of depression, given the higher rate of sleep disturbances in the elderly (Fragoso & Gill, 2007). Consider also an item that identifies cognitive problems. On the one hand, older adults may be less likely to report relatively minor cognitive changes. Older adults do not contend with the same daily cognitive demands as younger adults and therefore may be less likely to notice slight changes, whereas younger adults may notice these slight changes more readily because they interfere more significantly in their lives. Alternatively, older adults are at increased risk for cognitive difficulties and therefore may be more likely than younger adults to report these difficulties.

We do not know any studies that have directly examined whether the DSM depression items work equivalently well across younger and older age groups across the spectrum of depression severity. The accepted belief is that the psychometric properties of somatic and cognitive depression items are not equivalent across age groups (e.g., Brink et al., 1982; Yesavage et al., 1983). Specifically, it is thought that although the somatic and cognitive items may be quite discriminating in younger-adult depression, they may be less so in older adult depression. One could begin to investigate this idea using stacked factor analytic invariance models across younger and older age groups. Although these models could show that

*Corresponding author. Email: balsis@tamu.edu

ISSN 1360–7863 print/ISSN 1364–6915 online
© 2008 Taylor & Francis
DOI: 10.1080/13607860802428000
http://www.informaworld.com
the interrelationships among depression items differ across groups, they could not precisely show whether the items function differently at varying levels of depression. That is, they could not test whether a particular item is more or less likely to be endorsed by older as compared to younger adults with equivalent levels of depression. Item response theory (IRT)-based differential item functioning (DIF) analyses can accomplish this goal. These analyses can reveal in a much more detailed way the metric (non)equivalence of depression items across age groups. The objective of this analysis is to determine whether older and younger adults with equivalent levels of latent depression are equally likely to endorse particular DSM diagnostic items. The specifications of these analyses are outlined in the ‘Method section’.

We predicted that the analyses would reveal several different patterns. We anticipated that older adults would be more likely to endorse somatic items of depression than younger adults with similar levels of depression. Of particular interest to us was the suicide item. Although older adults are more likely to have completed a suicide attempt, younger adults are more likely to endorse suicidal thoughts and behaviors (Flint, Hays, Krishnan, Meador, & Blazer, 1998). Because the focus of the DSM suicide item is on behavior, ideation, and unsuccessful attempts, we predicted that older adults would be less likely to endorse this item compared with similarly depressed younger adults. We were not certain what to anticipate for the cognitive item, so we examined whether older adults would be more or less likely to endorse it compared with their younger counterparts. In contrast to our predictions regarding the somatic, cognitive, and suicide items, we had no strong predictions about the worthlessness/guilt or the fatigue items but anticipated that they would contain little or no measurement artifact. That is, we expected that they would be endorsed at similar rates by younger and older adults. To test these hypotheses, we analyzed data from participants who participated in the National Epidemiological Study on Alcoholism and Related Conditions (NESARC).

Method

Participants and procedure

The data used in this analysis come from the 2000 to 2001 wave of the NESARC study conducted by the National Institute of Alcohol Abuse and Alcoholism. This dataset consists of an epidemiological sample of noninstitutionalized adults 18–98 years old in the United States. Particular populations were oversampled in comparison with the US Census Bureau demographic data. Women and men were sampled at a rate equal to their proportion in the United States, and this strategy naturally resulted in more women being interviewed, particularly at older ages (Grant, Kaplan, Shephard, & Moore, 2003). Patients were not screened for dementia, but participants who appeared mentally or physically incapable of completing the survey were excluded at the discretion of the interviewer. Data from the 43,093 participants (Grant et al., 2003; Grant, Stinson, Dawson, Chou, & Ruan, 2005) were collected by 1800 trained interviewers.

The analyses for the current study included only younger (age 18–34 years) and older (age 65–98 years) participants who responded positively to at least one of two depression screener items and also responded to all relevant subsequent depression questions (11% of the participants who responded positively to at least one of the two screener items had missing data on at least one of the subsequent depression items). In the end, subsamples of 3734 younger adults and 1808 older adults were retained for the analyses (see Table 1 for participant characteristics). Both groups reported on their most recent bout of depression. For younger adults, this occurred, on average, at age 24 years. For older adults, it occurred, on average, at age 62 years.

Materials

Depressive symptoms were assessed with a structured diagnostic interview based on DSM-IV features, the Alcohol Use Disorder and Associated Disabilities Interview Schedule – DSM-IV version (AUDADIS-IV). Among other items, the interview contained a series of yes/no questions regarding

<table>
<thead>
<tr>
<th>Table 1. Demographics.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Younger (age 18–34 years)</strong></td>
</tr>
<tr>
<td><strong>n = 3734</strong></td>
</tr>
<tr>
<td>Age</td>
</tr>
<tr>
<td>Age of depressive symptoms</td>
</tr>
<tr>
<td>Gender</td>
</tr>
<tr>
<td>White</td>
</tr>
<tr>
<td>Black</td>
</tr>
<tr>
<td>Asian</td>
</tr>
<tr>
<td>American Indian</td>
</tr>
<tr>
<td>Native Hawaiian</td>
</tr>
<tr>
<td>Hispanic</td>
</tr>
<tr>
<td>White</td>
</tr>
<tr>
<td>Black</td>
</tr>
<tr>
<td>Asian</td>
</tr>
<tr>
<td>American Indian</td>
</tr>
<tr>
<td>Native Hawaiian</td>
</tr>
<tr>
<td>Hispanic</td>
</tr>
</tbody>
</table>

Note: SD = Standard deviation.
DSM-IV depression features. Nine diagnostic features make up the DSM-IV depression item set. Two of these features (#s 1 and 2) serve as screening criteria. That is, to meet a diagnosis of depression, an individual has to endorse either feature 1 (low mood) or feature 2 (anhedonia). Items based on these two features were used as screening questions for the interview (Appendix). First, interviewers asked participants if they ever had a 2-week period when they had a low mood or had trouble initiating behavior. If the respondent endorsed either of these items, the interviewer asked whether the subsequent depression items occurred simultaneously with the screening symptom(s). If the respondent did not endorse either of these screening items, the interviewer did not ask the subsequent depression items. For the purposes of the present study, we analyzed data from the participants’ most recent depressive episodes and determined whether a participant met each of the seven additional DSM-IV depression features by combining questions from the interview (the Appendix shows more specifically how questions from the interview mapped onto DSM-IV diagnostic features). These seven features were the focus of the present analyses.

Participants were allowed to skip questions if they so chose or could not give a yes/no answer, but only data from participants who answered all relevant questions were included in the present analyses. The internal consistency of all seven items was relatively high (\(a = 0.75\) for younger adults and 0.77 for older adults).

**Data analysis**

Within an IRT framework, the analyses examined measurement bias between younger and older age groups in seven of the DSM-IV depression diagnostic features. Measurement bias exists when younger and older adults who are matched with respect to their latent level of depression endorse a different number of features, on average.

The probability of endorsing each item at each response option level (does not apply, applies) was predicted by fitting a two parameter logistic model to observed data. A discrimination (\(a\)) and threshold (\(b\)) parameter were estimated for each item. The \(a\) parameter indicates how related an item is to the latent depression continuum, and the \(b\) parameters determine where on the latent continuum each item discriminates optimally. Together, the \(a\) and \(b\) parameters define the probabilities that each item is endorsed at all levels of the latent variable. The latent variable, \(\theta\), represents latent depression. It is given a standard scale (i.e., mean = 0, variance = 1). Response probabilities (0–100%) for each item are plotted at 201 levels of \(\theta\). The 201 levels of \(\theta\) range between –4.00 and +4.00 standard deviations (SDs) and are marked off at intervals of size 0.04 (–4.00, –3.96, –3.92, . . . , 3.92, 3.96, 4.00).

To establish the \(a\) and \(b\) parameters for each item, IRT-based likelihood-ratio DIF (IRT-LR-DIF) estimation (Thissen, Steinberg, & Gerrard, 1986) was implemented using the IRT-LR-DIF program (v.2.0b, Thissen, 2002). For each item, a model with item parameters restricted so as to be equal for younger and older adults is compared with a model that permits item parameters to vary between groups.

Each item for each age group has an item characteristic curve. For each age group, the seven-item characteristic curves were combined together to create a test characteristic curve. For example, at 1.25 SDs of depression, younger adults may have a 1% probability of endorsing item 1, a 9% probability of endorsing item 2, a 75% probability of endorsing item 3, a 54% probability of endorsing item 4, and so on. These values were combined by summing at every given level of \(\theta\) each of these probabilities. The resulting value, at 1.25 SDs = 3.45, indicates that at this particular level of depression a younger adult is expected to endorse 3.45 features on average. This additive procedure was repeated at all 201 \(\theta\) values. Then the resultant scores were plotted in a figure in which the abscissa ranged from –4.00 to +4.00, and the ordinate reflected the range of response options (in this case, 0–7).

This procedure was conducted separately for the younger- and older-adult curves; then both sets of curves were plotted together for comparison purposes. This is appropriate, as all have been set to the same scale when the \(a\) and \(b\) parameters were estimated.

**Results**

In these subsamples, younger and older adults had a relatively high likelihood of endorsing each screening item (Table 2). Ninety-two percent of both younger and older adults endorsed the low mood item. Meanwhile 77% of younger adults and 68% of older adults endorsed the anhedonia item. The difference between younger and older adults for the anhedonia item was statistically significant, \(\chi^2 = 48.02, p < 0.001\), with younger adults more likely to endorse it. Recall again that these items were screening items, so either item had to be endorsed before a participant was presented with the additional items. As a result, a high rate of endorsement is expected for each item across these particular subgroups.

For the additional seven items, again, both younger and older adults had a high likelihood of endorsement. Younger adults were more likely to endorse the psychomotor agitation/retardation item (43% endorsement). In comparison with other items, the younger group was most likely to endorse the sleep-change item (76%). Older adults were least likely to endorse the suicide item (31%) and most likely to endorse the sleep change item (65%). Although the rate at which these items were endorsed varied within groups, across the subsamples, younger adults were always more likely to
endorse each of these seven items. $\chi^2$ difference tests ranged from 46.52, $p < 0.001$, for the weight/appetite-change item to 310.08, $p < 0.001$, for the concentration item.

Subsequent IRT-based analyses revealed systematic patterns (Figure 1). Notice that all of the item characteristic curves increase progressively. This finding suggests that participants (younger or older) with low levels of depression were relatively less likely to endorse each item. At the same time, participants with high levels of depression were more likely to endorse each item. This same basic pattern held across all seven items. Greater levels of depression were associated with a higher likelihood of endorsement.

Next, consider the measurement bias uncovered across the items. Older adults were more likely to

Table 2. Prevalence of depressive symptoms in younger and older subsamples.

<table>
<thead>
<tr>
<th></th>
<th>Younger (age 18–34 years)</th>
<th>Older (age 65–98 years)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Low mood</td>
<td>3430 (92%)</td>
<td>1667 (92%)</td>
</tr>
<tr>
<td>2. Anhedonia*</td>
<td>2864 (77%)</td>
<td>1229 (68%)</td>
</tr>
<tr>
<td>3. Weight/appetite change*</td>
<td>2516 (67%)</td>
<td>1049 (58%)</td>
</tr>
<tr>
<td>4. Sleep change*</td>
<td>2846 (76%)</td>
<td>1166 (65%)</td>
</tr>
<tr>
<td>5. Psychomotor agitation/retardation*</td>
<td>1616 (43%)</td>
<td>602 (33%)</td>
</tr>
<tr>
<td>6. Fatigue*</td>
<td>2407 (65%)</td>
<td>903 (50%)</td>
</tr>
<tr>
<td>7. Worthlessness/guilt*</td>
<td>2246 (60%)</td>
<td>679 (38%)</td>
</tr>
<tr>
<td>8. Diminished concentration*</td>
<td>2778 (74%)</td>
<td>915 (51%)</td>
</tr>
<tr>
<td>9. Death/suicidal thoughts*</td>
<td>1774 (48%)</td>
<td>552 (31%)</td>
</tr>
<tr>
<td>Total # items endorsed*</td>
<td>6.02 (SD = 2.34)</td>
<td>4.85 (SD = 2.43)</td>
</tr>
</tbody>
</table>

Note: *Indicates that $\chi^2$ difference test was significant beyond $p < 0.001$.

Figure 1. Item and test characteristic curves for DSM-IV depression criteria.

Notes: - - = Older adults. — = Younger adults. # = No. of criteria. *Indicates that $\chi^2$ difference test was significant beyond $p = 0.05$. 
endorse somatic items than younger adults with the same level of latent depression. The weight-/appetite-change, sleep-change, psychomotor-retardation/agitation, and fatigue items were all more likely to be endorsed by older adults than younger adults with the same latent level of depression. For example, at 0.5 SDs of depression (in this very depressed sample), older adults had an 86% chance of endorsing the psychomotor/retardation item; whereas only 67% of younger adults with that same level of depression were likely to endorse that item. This same general pattern held for the other somatic items.

At the same time, older adults with the same latent depression (relative to their younger-adult counterparts) were less likely to endorse the cognitive and suicide item. In particular, older adults with 0.0 SDs of depression had a 90% likelihood of endorsing the cognitive item; whereas 98% of younger adults had that same likelihood. Similarly, older adults were much less likely to endorse suicidal ideation and behavior than younger adults with the same level of latent depression. Sixty percent of older adults with 1.0 SD of depression were expected to endorse this item. Meanwhile, 80% of younger adults with the same level of depression were expected to endorse it. Older and younger adults were similarly likely to endorse the mood (worthlessness/guilt) item. As an example, both older and younger adults with 0.5 SDs of latent depression had approximately a 90% likelihood of endorsing it.

A second noteworthy pattern was that when these items were combined to form the test characteristic curve encompassing all seven depression items (Figure 1), the direction and magnitude of the measurement artifact largely cancelled itself out. This is shown pictorially in that the curves that define the scale for each age group overlapped to a large extent. Notice one example of the overlap; younger and older adults with a mean level (0.0 SDs) of depression (in this very depressed sample) were both expected to endorse approximately five items. The large overlap between the curves at this level is noteworthy because it suggests that the additive effects of the measurement artifact at the item level cancel themselves out at the scale level.

**Discussion**

Overall, depression diagnostic features were endorsed at a higher rate by younger adults than older adults. This is consistent with existing literature, which shows that the rate of depression is much higher in younger adults than it is in older adults (e.g., Narrow, Rae, Robins, & Regier, 2002). Of particular note, this pattern was maintained for each non-screener item. Younger adults consistently were more likely to endorse each depression criterion.

Although each item was more likely to be endorsed in general by younger adults, IRT analyses explored the likelihood they would be endorsed when controlling for level of latent depression across groups. The IRT analyses revealed that older adults were more likely to endorse somatic items and less likely to endorse cognitive items when compared with equally depressed younger adults. This is consistent with the long-held belief that older adults, regardless of their level of depression, have more sleep problems, weight change, and general health problems. This finding also is consistent with the widely held belief that older adults have fewer cognitive demands to contend with than younger adults.

Of particular note, older adults were much less likely to endorse the suicide item than similarly depressed younger adults. This finding has several potential explanations. First, it may indicate that older adults actually experience less suicidal ideation than similarly depressed younger adults. Alternatively, this finding could be the representative of systematic underreporting of suicidal ideation by older adults. Given that older adults, particularly White men, are more likely to commit suicide than younger adults, clinicians may need to rely on multiple lines of inquiry and evaluation to assess suicide risk in these older adults.

The other noteworthy finding was that the measurement artifact within each item cancelled itself out at the scale level. That is, when the overreporting (relative to younger adults) of the somatic symptoms and the underreporting of the cognitive and suicide symptoms were considered together, the direction and magnitude of the effects largely cancelled each other out. This findings suggests that the DSM diagnostic items, when considered together, work equally well across age groups. However, our data also suggest that the measurement of the particular symptoms of depression may be significantly compromised, with older adults being more likely to report somatic symptoms and younger adults being more likely to report cognitive and suicide symptoms. These results are most relevant for clinicians and researchers who attempt to identify and compare individual depressive features across age groups.

This study has several limitations, including its cross-sectional design. Although the reported patterns seem to fit with the age-associated explanations described above, we cannot be certain whether some of the effects occurred as a result of cohort-associated factors. Future examinations of longitudinal and cross-sequential data may help confirm that these findings are in fact age-associated (and not cohort associated). Note also that older adults in this sample were reporting on depressive symptoms that occurred on average at age 62 years. If the older adults were reporting of depressive symptoms that occurred at an older age, the findings might be magnified substantially. Future studies may explore whether the items function differently when participants report only current symptoms. Another noteworthy point is that it is not clear if these symptoms are associated with
some functional impairment or are due to a medical condition or medication. One issue to consider is that the cause or consequence of certain symptoms in a particular individual can be very difficult to establish. For this reason, we chose simply to focus on the DSM symptoms and not their causes or consequences. Other limitations of the study have to do with the subsamples chosen for the analysis. Among other issues, only data from those who answered positively to the screener questions were analyzed, but the items themselves may function differently for those who answered positively and those who answered negatively. These analyses only generalize to those who would have answered positively to these screening items. Future investigations might determine whether the items perform differently across these groups. It is also important to note that only data from those who provided responses to all of these depression questions were analyzed. As such, these findings also generalize only to individuals who would provide complete data. A final point to consider is that these analyses examine only the DSM depression symptoms and not other aspects of the major depressive definitional requirements.

References


Table A1. Algorithm used to create DSM-IV major depressive episode diagnostic features.

<table>
<thead>
<tr>
<th>DSM-IV features (paraphrased)</th>
<th>NESARC items (paraphrased)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Low mood</td>
<td>1. Felt sad, blue, depressed, or down</td>
</tr>
<tr>
<td>2. Anhedonia</td>
<td>2. Didn’t care about the things you typically cared about or didn’t enjoy the things you typically enjoyed.</td>
</tr>
</tbody>
</table>
| 3. Appetite/weight change<sup>a</sup> | 3a. Lost 2+ pounds a week for several weeks  
3b. Lost appetite  
3c. Gained 2+ pounds a week for several weeks  
3d. Gained appetite |
| 4. Sleep problems<sup>a</sup> | 4a. Had trouble falling asleep  
4b. Had trouble staying asleep  
4c. Slept more than usual |
| 5. Psychomotor agitation/retardation<sup>a</sup> | 5a. Moved or talked slower than usual  
5b. Fidgeted or paced most of the time |
| 6. Fatigue                     | 6. Felt tired, or got tired more easily |
| 7. Worthlessness/Guilt<sup>a</sup> | 7a. Felt worthless  
7b. Felt guilty |
| 8. Poor concentration/poor thinking/poor decision making<sup>a</sup> | 8a. Had trouble concentrating/staying focused  
8b. Had trouble making decisions |
| 9. Death thoughts/suicidal thoughts or behaviors<sup>a</sup> | 9a. Attempted suicide  
9b. Thought about suicide  
9c. Wanted to die  
9d. Thought about death a lot |

Notes: <sup>a</sup>Indicates that item requires endorsement of at least one of the associated NESARC items. NESARC = National Epidemiological Study on Alcoholism and Related Conditions.