Correspondence of Verbal Descriptor and Numeric Rating Scales for Pain Intensity: An Item Response Theory Calibration

Maria Orlando Edelen1 and Debra Saliba1,2,3

1Health Unit, RAND Corporation, Santa Monica, CA.
2VA Greater Los Angeles GRECC and HSR&D Center of Excellence, Los Angeles, CA.
3UCLA/Los Angeles Jewish Homes Borun Center, Los Angeles, CA.

Address correspondence to Maria Orlando Edelen, PhD, RAND Corporation, Santa Monica, CA. Email: orlando@rand.org

Background. Assessing pain intensity in older adults is critical and challenging. There is debate about the most effective way to ask older adults to describe their pain severity, and clinicians vary in their preferred approaches, making comparison of pain intensity scores across settings difficult.

Methods. A total of 3,676 residents from 71 community nursing homes across eight states were asked about pain presence. The 1,960 residents who reported pain within the past 5 days (53% of total, 70% female; age: $M = 77.9, SD = 12.4$) were included in analyses. Those who reported pain were also asked to provide a rating of pain intensity using either a verbal descriptor scale (VDS; mild, moderate, severe, and very severe and horrible), a numeric rating scale (NRS; 0 = no pain to 10 = worst pain imaginable), or both. We used item response theory (IRT) methods to identify the correspondence between the VDS and the NRS response options by estimating item parameters for these and five additional pain items.

Results. The sample reported moderate amounts of pain on average. Examination of the IRT location parameters for the pain intensity items indicated the following approximate correspondence: VDS mild $≈$ NRS 1–4, VDS moderate $≈$ NRS 5–7, VDS severe $≈$ NRS 8–9, and VDS very severe, horrible $≈$ NRS 10.

Conclusion. This IRT calibration provides a crosswalk between the two response scales so that either can be used in practice depending on the preference of the clinician and respondent.

Key Words: Pain—IRT—Measurement.

Received September 18, 2009; Accepted December 8, 2009

Decision Editor: Luigi Ferrucci, MD, PhD

Pain is a prevalent experience among older adults (1,2) and adds to depression, social isolation, and functional disability (3–8). Despite evidence that the treatment of pain can increase quality of life (9), there is general agreement that pain is underdetected and understudied among older adults (1,10–16). Although the optimal method to detect pain is via self-report (6,17), detection of pain via self-report can be difficult in this population for a variety of reasons. Older adults may not spontaneously report their pain because they attribute it to the aging process (18), and a significant amount of pain goes undetected if assessment is based on observation alone (19,20). For these reasons, current guidelines suggest the routine use of self-report pain measures to increase pain detection in elderly adults (21–24).

There is considerable debate about the most effective way to ask older adults, particularly those who are frail or institutionalized, to describe their pain. Two of the most common and psychometrically supported methods are the 0–10 numeric rating scale (NRS) (25) and verbal descriptor scales (VDS) (24). The NRS typically asks the respondent to rate their pain on a scale of 0–10 with 0 being no pain and 10 being the worst pain possible. A VDS typically asks residents to choose a verbal description of their pain (eg, mild, moderate, severe) on a 4- or 5-point scale.

There is some evidence that the VDS performs better than the NRS with respect to respondent preference, completion rates, sensitivity, and reliability (20,26–28), but the evidence is not conclusive and many argue strongly in support of NRS (23,29), which has demonstrated reliability and validity (30), and is the more commonly used intensity scale in the general population (31,32). The lack of consensus on the relative performance of these two methods and the fact that clinicians and patients vary in their preferred approaches (20) has resulted in routine use of both types of scales in practice, making comparison of pain intensity scores across settings difficult (33).

Because these two types of scales are routinely used in practice, it would be useful to devise a way to calibrate pain levels accurately across the scales (20). In addition, a crosswalk or calibration would allow the inclusion of both scales in the planned revision of the Minimum Data Set (MDS).
VDS AND NRS FOR PAIN INTENSITY

3.0), giving providers and residents a choice between the severity scales. However, although several studies have compared pain intensity scales with respect to response rates, preference, and reliability, there have been few attempts to create a common metric for the different scales (34). Jones and colleagues (34) reviewed the literature and summarized efforts to categorize the 1–10 NRS onto pain intensity categorizations of mild, moderate, and severe. The mild category most often had NRS values of 1–4, but some researchers excluded 4 or included 5; moderate assignments were 4–6, 5–6, 5–7, 6–7, or 6–8; and severe included 7–10, 8–10, or 9–10. Jones and colleagues then collapsed a 6-point VDS scale into the three intensity categories and provided a correspondence to the 1–10 NRS based on data from 270 nursing home residents as follows: 1–4 = mild, 5–6 = moderate, and 7–10 = severe.

The work presented by Jones and colleagues (34) provides insights into the likely correspondence between the NRS and VDS categories. However, the cutoff points vary across studies, and the methods to derive the cut points are largely not empirically based. Even the correspondence provided by Jones and colleagues (34) was derived primarily by qualitative methods and validated using simple measures of correspondence and a relatively small sample.

In this article, we aim to improve upon this work by bringing modern measurement theory to bear on this question. Specifically, we use item response theory (IRT) (35) to provide a cross-calibration of the 4-point VDS and the 0–10 NRS among a large national sample of nursing home residents whose pain was assessed as part of the national field trial of the proposed MDS 3.0. IRT is a measurement approach that characterizes the relationship between a latent construct (in this case pain experience) and a set of items measuring that construct (eg. pain intensity questions). IRT has been applied in countless measurement settings and has been useful in studying responses of older adults to assessments of functional status, depression, cognitive status, and other constructs (36–39).

The IRT approach has several advantages (40): The results from an IRT calibration are highly generalizable beyond the sample they were derived from assuming that sample is representative. IRT uses the relationship of items to a common underlying construct to position all item responses on a single continuum so they can be compared, IRT does not need to assume equal intervals between response categories, and the results of an IRT calibration can be used to generate graphical representations of item responses that can be very useful for examining item correspondence.

IRT is particularly useful for examining the correspondence between NRS and VDS responses in this large geographically diverse sample given that several respondents answered both the VDS and the NRS items. Furthermore, respondents were asked additional pain items, and responses to these items can be combined with the intensity items to represent a strong underlying construct of resident’s perceived pain experience. The inclusion of these additional items lends stability to the cross-calibration and increases precision of the item parameter estimates.

METHODS

Participants

Study participants took part in the national MDS 3.0 field trial, a trial involving residents from 71 community nursing homes in 8 states. The trial was designed to test the performance of proposed revisions to the pain items in the MDS and consisted of residents scheduled for routine MDS 2.0 assessments who were then assigned based on a preset algorithm to (a) a crosswalk sample (N = 3,258) that received the trial MDS 3.0 items and was intended to provide direct comparisons to between MDS 3.0 and 2.0 assessments or (b) a validation sample (N = 418) that received an augmented set of MDS 3.0 items as well as gold-standard measures for constructs where available to examine the validity of the MDS 3.0 items.

As part of the MDS 3.0 test battery, all 3,676 residents were asked whether they had experienced any pain or hurting in the past 5 days. Just over half of the residents (54%, n = 1,971) responded yes to this question, 35% (n = 1,290) reported no pain, and 11% (n = 415) either were not asked the question because of inability to be interviewed (n = 307) or provided nonresponsive answers (n = 108). Among those who answered yes to the question about pain presence, 11 respondents had missing values for all the remaining pain questions. Thus, the analyses reported in this article are based on a sample of n = 1,960 (see Table 1).

Measures

All respondents who endorsed the initial pain screening question (“Have you had pain or hurting at any time in the last 5 days? Y/N”) were asked a series of questions designed to characterize the pain experience.
Pain intensity was measured with two approaches. The first approach asked the respondent to “Please rate the intensity of your worst pain over the last 5 days” and respondents were read and shown a 4-point VDS: mild, moderate, severe, and very severe/horrible. The second approach used a NRS and asked residents: “Please rate your worst pain over the last 5 days on a 0 to 10 scale with zero being no pain at all and ten as the worst pain you can imagine.” These respondents were read and shown a 0–10 scale with 0 anchored to no pain and 10 anchored to worst pain you can imagine. The design for inclusion of these items varied in the crosswalk and validation samples (see Table 1). All 242 residents who reported pain in the validation sample were presented with both versions; the NRS was embedded in the pain assessment and the VDS was presented later in the survey after questions on other topics. In the crosswalk sample, data collectors in each facility were instructed to use the scale they typically used or that the resident had used on prior assessments. In this sample, nurse collectors were instructed to only ask a second intensity scale if the resident had difficulty with the first one that was attempted. However, some nurses appear to have presented both scales regardless of difficulty. A response of 0 for the NRS item was considered inconsistent given the endorsement of pain in the last 5 days and was set to missing for the 10 respondents who chose it. Additionally, because of the low frequency of responses in NRS Category 1 (n = 10), it was combined with Category 2 of the NRS for analyses.

In addition to intensity, respondents in both samples were also asked to rate the frequency of their pain on a 4-point scale (1 = almost constantly to 4 = never). Four dichotomous (Y/N) items assessing “functional limitations due to pain” were also queried, two were asked of both the crosswalk and validation samples (pain makes it hard to sleep and pain has limited your activities) and two were included only in the validation sample protocol (pain makes it hard to get out of bed and pain makes it hard to spend time with others). These functional items have been pilot tested and shown to provide additional information about pain beyond that captured by pain intensity items (41). Finally, the validation sample also responded to an item assessing the degree of bother associated with the pain on a 4-point scale (1 = not at all to 4 = a great deal). Only one resident endorsed the “not at all” option for this item; thus, “not at all” was combined with the adjacent category “a little” for analyses.

Analytic Approach

The goal of these analyses was to identify the correspondence between the two intensity scales’ response options by considering their relative relationship to the resident’s underlying pain experience. We applied IRT methods to accomplish this.

Preliminary analyses.—We first generated descriptive statistics for the eight pain items. Because IRT seeks to relate each item to a common underlying construct, it is necessary to test whether the combined items form a unidimensional construct. We therefore conducted confirmatory factor analysis with the Mplus software (42) to ensure the set of items was sufficiently unidimensional for IRT analysis. This assumption is critical to the clear interpretation of item parameters. Because of IRT’s built in linking mechanism (43), items that are completed by only a subsample of respondents can be included in a calibration provided that there are some common items that can serve to link the full sample of respondents. In this application, all respondents completed three items (pain frequency and two of the limitations items) and 16% of respondents completed both NRS and VDS intensity items. This extent of common items across respondents allows for the simultaneous calibration of all eight pain items.

Item calibration and evaluation of model results.—We used the graded response model of Samejima (44) implemented with MULTILOG (45) to estimate item parameters for the eight pain items. Like other unidimensional IRT models, the graded model specifies the relationship between an individual’s observed responses to a set of items and the unobserved (or latent) trait that is being measured by the item set, in this case pain experience. One result of an IRT calibration is a set of continuous item characteristic curves (ICCs), most commonly defined as logistic functions, which describe the probability of endorsing each response category of an item given the scale value of the measured construct.

For items with multiple ordered response categories like the pain intensity, frequency, and bother items, each ICC of graded model specifies the probability of choosing a response as a function of one item slope parameter (a) and a location (b) parameter for each response category transition. For example, for items with three response categories, two location parameters (b1, b2) are estimated, whereas for items with two response categories, a single location parameter (b1) is estimated. The slope reflects the strength of the relationship of the item to the underlying construct being measured, analogous to an item loading in factor analysis. The location parameters reflect the spacing of the item responses across the construct continuum, analogous to item thresholds in factor analysis with mean structure.

We examined the overall results of the IRT model including evaluation of item fit, and examination of the item parameters. Item fit was evaluated using Bjoerner and colleagues (46) extension of S–X2 (47,48).

Mapping NRS and VDS.—After examining the overall results of the IRT models, including evaluation of model fit, we considered several aspects of the IRT calibration results to identify the optimal mapping of NRS responses to VDS categories. Because the IRT calibration places all the item responses on the same underlying continuum, it is possible...
to create plots that display multiple items and show how the response options line up. Thus, we first examined plots of the ICCs to display the overlap in response categories between the VDS and NRS. Because items and scores are placed on the same underlying continuum, another way to use the IRT analytic results to explore correspondence of item response options is to examine the expected scores associated with each response option. Therefore, our second step was to calculate the model-based expected pain experience score associated with each response option for the two items and examine their correspondence in order to understand how the responses line up numerically.

These preliminary examinations provided a good indication of the mapping for many of the NRS responses into VDS categories, but there was still some question of the best correspondence for NRS response 7. A final set of plots, one for each of VDS response options moderate and severe, allowed us to evaluate the relative suitability of the placement of NRS response option 7. Each of these plots contained one VDS response option and a curve representing the sum of selected NRS response option densities. This allowed us to determine whether the inclusion of the NRS option 7 in VDS moderate created a closer match than when it was excluded and considered in VDS severe instead.

RESULTS

Preliminary Analyses

Table 2 provides descriptive information for the 1,960 residents in the analytic sample and also displays the number of respondents and response frequencies for each of the eight pain items. Nearly all residents in the sample answered the pain frequency question and two of the items querying limitations due to pain (hard to sleep and limited activities).

With regard to the pain intensity items, n = 298 responded to both VDS and NRS pain frequency items, n = 804 responded only to NRS, and n = 823 responded only to VDS. Responses to the remaining three items (bother, hard to get out of bed, and hard to be with others) were restricted to those in the validation sample.

The response frequencies indicate a sample of patients with moderate amounts of pain. For intensity, residents tended to endorse responses 5–8 on the NRS, and moderate was the most frequent response on the VDS. The middle response options were most frequently endorsed for pain frequency (frequently and occasionally) and bother (a moderate amount); nearly half the sample reported that pain limited activities and made it hard to get out of bed. The remaining two limitations were endorsed at lower rates.

A confirmatory factor analysis was conducted using Mplus (41) to ensure that the set of items was sufficiently unidimensional for IRT analysis. The one factor solution fit the data quite well, $\chi^2(n = 1,960, 17) = 62.144$, comparative fit index = .985, non-normed fit index = .986, root mean square error of approximation = .037), providing support for the applicability of the data for IRT analysis.

Table 2. Sample Demographics and Response Frequencies (percent of responders) for Pain Items (N = 1,960)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Age, M (SD)</th>
<th>Female, %</th>
<th>Ethnicity, %*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>77.87 (12.42)</td>
<td>69.6</td>
<td>White 85.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>African American 7.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Hispanic/Latino 2.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Asian 0.77</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Other 4.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Primary language English, % 99.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Education, ≥ high school, % 74.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Admission MDS, % 59.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>MDS-ADL hierarchy score (range 0–6), M (SD) 3.29 (1.47)</td>
</tr>
<tr>
<td>Pain intensity, VDS (n = 1,121)†</td>
<td></td>
<td></td>
<td>Gentle 0.38</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Moderate 4.79</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Severe 6.44</td>
</tr>
<tr>
<td>Pain bother (n = 236)†</td>
<td></td>
<td></td>
<td>Rarely 6.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Occasionally 38.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Frequently 33.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Almost constantly 21.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Not at all/a little 25.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>A moderate amount 45.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>A great deal 28.8</td>
</tr>
<tr>
<td>Limitations due to pain</td>
<td>No</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>Hard to sleep (n = 1,954)</td>
<td>64.2</td>
<td>35.8</td>
<td></td>
</tr>
<tr>
<td>Limited activities (n = 1,939)</td>
<td>53.9</td>
<td>46.1</td>
<td></td>
</tr>
<tr>
<td>Hard to get out of bed (n = 237)</td>
<td>51.9</td>
<td>48.1</td>
<td></td>
</tr>
<tr>
<td>Hard to spend time with others (n = 235)</td>
<td>78.3</td>
<td>21.7</td>
<td></td>
</tr>
</tbody>
</table>

Notes: *Information on race/ethnicity was only asked for new admissions. This percent is based on only 67% of the sample (n = 1,306).
†n = 306 responded to both VDS and NRS pain frequency items, n = 807 responded only to NRS, and n = 816 responded only to VDS. ADL = Activities of Daily Living; MDS = Minimum Data Set; NRS = numeric rating scale; VDS = verbal descriptor scale.

The graded model estimation converged successfully. With the exception of the NRS item, which could not be evaluated for fit because it exceeded the maximum response options allowed by the software (45), all items displayed acceptable fit, indicating that the IRT model is an adequate representation of the data. Table 3 displays the item parameter
estimates and their standard errors for each of the pain items based on the graded model calibration. Slope estimates \( (a) \) are all strong and positive (range = 1.19–2.80), indicating that all items are strong indicators of the latent construct (pain experience). The two pain intensity items and the bother item had the highest slopes. The threshold parameters ranged from −2.23 to 1.55, indicating a good representation of varying levels of the pain experience continuum.

**Mapping NRS and VDS**

Figure 1 displays the item category trace lines for the VDS and NRS items. In these plots, the x-axis represents the underlying continuum of pain experience measured by the item set and the y-axis represents the probability of response option endorsement. For the VDS item, the response category trace lines are quite distinct; each response option covers some unique area along the pain experience continuum and it is fairly easy to discern which option is most likely for a given level of pain experience (except, of course, at the crossing points). For example, a person with a pain experience score of −0.5 is extremely likely to endorse response option moderate and has only a small probability of endorsing response options mild or severe. The NRS item does not have these properties. The response options overlap a great deal, and response options 4, 6, and 9 do not cover any unique area (ie, these category trace lines are completely contained within other response options). To continue the example, a person with a pain experience score of −0.5 is most likely to endorse NRS option 5, but also has nonzero probabilities of endorsing all the other response options, as the curves for options 1–2, 3, 4, 6, 7, 8, 9, and 10 all have shared area with the curve for option 5. This pattern implies that the 10 categories do not provide distinct information; for example, a person who responds 4 has a very similar pain experience to someone who responds 3 or 5.

Although comparing the VDS lines to the NRS lines in Figure 1 provides an initial idea of the correspondence between these two items’ response options, the best mapping is not obvious from this approach. Another perspective is gained by examining the model-based expected pain experience score associated with each response option of the two items. Figure 2 displays this correspondence. As in Figure 1, the x-axis represents the underlying pain experience continuum, but the scale range is narrower because all scores fall within the range from −2.5 to 2.5. In Figure 2, the VDS response option scores are displayed with error lines representing ±1 SE. From this display, most NRS options clearly fall within a VDS category. NRS responses 1–2, 3, and 4 are within mild; 5 and 6 are within moderate; 8 and 9 are within severe; and 10 is within horrible. However, the score corresponding to NRS response 7 falls within 1 SE of two different VDS categories (moderate and severe) and therefore cannot be cleanly assigned to a single VDS category based solely on this analysis.

A final approach informed the mapping for the remaining NRS response by constructing plots for each of the two VDS response options, moderate and severe, which also displayed the sum of selected NRS response option densities. Figure 3 shows these results displaying VDS moderate with NRS 5–6 and 5–7 and VDS severe with NRS 7–9 and 8–9. The optimal placement of NRS response option 7 is

**Table 3. Item Parameter Estimates and Standard Errors for Eight Items**

<table>
<thead>
<tr>
<th>Item</th>
<th>( a )</th>
<th>( b_1 )</th>
<th>( b_2 )</th>
<th>( b_3 )</th>
<th>( b_4 )</th>
<th>( b_5 )</th>
<th>( b_6 )</th>
<th>( b_7 )</th>
<th>( b_8 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intensity, NRS</td>
<td>2.09(0.10)</td>
<td>−2.83(0.17)</td>
<td>−2.06(0.11)</td>
<td>−1.47(0.08)</td>
<td>−0.99(0.06)</td>
<td>−0.29(0.05)</td>
<td>0.15(0.05)</td>
<td>0.66(0.05)</td>
<td>1.22(0.06)</td>
</tr>
<tr>
<td>Intensity, VDS</td>
<td>2.73(0.15)</td>
<td>−1.04(0.05)</td>
<td>0.44(0.04)</td>
<td>1.56(0.07)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Frequency</td>
<td>1.71(0.08)</td>
<td>−2.21(0.10)</td>
<td>−0.18(0.04)</td>
<td>1.12(0.05)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bother</td>
<td>2.20(0.28)</td>
<td>−0.89(0.13)</td>
<td>0.75(0.12)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hard to sleep</td>
<td>1.37(0.10)</td>
<td>0.58(0.06)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bother</td>
<td>2.20(0.28)</td>
<td>−0.89(0.13)</td>
<td>0.75(0.12)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Frequency</td>
<td>1.71(0.08)</td>
<td>−2.21(0.10)</td>
<td>−0.18(0.04)</td>
<td>1.12(0.05)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bother</td>
<td>2.20(0.28)</td>
<td>−0.89(0.13)</td>
<td>0.75(0.12)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hard to sleep</td>
<td>1.37(0.10)</td>
<td>0.58(0.06)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Limited activities</td>
<td>1.29(0.09)</td>
<td>0.17(0.05)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hard get out of bed</td>
<td>1.20(0.24)</td>
<td>0.10(0.16)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hard time with others</td>
<td>1.25(0.28)</td>
<td>1.36(0.22)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Note: NRS = numeric rating scale; VDS = verbal descriptor scale.*
VDS AND NRS FOR PAIN INTENSITY

still not definite from visual inspection of these plots, although it seems to favor including option 7 with moderate. A calculation of the area between the lines weighted by the normal density revealed that placing NRS response option 7 with moderate resulted in a total area between the lines of 0.213, whereas NRS 7 with severe resulted in a total area between the lines of 0.426. Thus, to minimize the area between the lines, NRS 7 should be included within VDS moderate. Figure 4 shows the final solution with the following VDS \( \approx \) NRS mapping: mild \( \approx 1–4 \), moderate \( \approx 5–7 \), severe \( \approx 8–9 \), and very severe/horrible \( \approx 10 \).

**DISCUSSION**

Current guidelines suggest that assessment of pain in residential care facilities should be done via resident self-report.
The relatively high slopes for the VDS and NRS items in the IRT results indicate that both are strong indicators of the underlying pain experience. This lends support for the use of either of these items as screening questions for pain assessment. However, the precision of these items is not sufficient to replace a comprehensive pain assessment. The reliability of the VDS and NRS items are highest in the middle range of the pain experience continuum (±1 SD around the mean of 0) but even there the standard error of the score estimates is about 0.6 standard units, which is quite large. This limited precision of the single items makes them less than ideal for making conclusive inferences about individuals. Furthermore, although the performance of the VDS appears slightly better than the NRS (eg, higher slope, better spacing of item responses), there are other important aspects of performance to consider before concluding that one scale is superior to the other. These factors include the stability of the item, its sensitivity to change over time, and its suitability for various subgroups (eg, according to native language or ethnicity).

The article has a number of strengths including its use of a large national sample, its inclusion of residents with a range of cognitive impairment, and its use of other pain items to obtain stable item parameter estimates. In addition, IRT is a state-of-the-art methodology that is ideally suited for this type of cross-calibration, allowing for the use of incomplete data so that the maximum number of respondents and items are included in the analysis. There are also limitations to note. The results are based on a sample of adults in residential care facilities. Although almost 60% are newly admitted, these results do not necessarily generalize to community-dwelling healthier populations. In addition, the performance of these items when translated or used with persons for whom English is not a primary language may differ.

Future research should extend this mapping to include other common pain assessment techniques, including the faces pain scale and the pain thermometer. It would also be of interest to determine whether this mapping applies to populations other than the one represented in this study. Finally, it is important to note that although intensity is an important and frequently determined pain measure, a full pain assessment requires a more multidimensional query.

**Funding**

This work was funded by the U.S. Department of Veterans Affairs, Veterans Health Administration, VA Health Services Research & Development Service (Project #SDR 03–217), and the Centers for Medicare and Medicaid Services. The views expressed in this article are those of the authors and do not necessarily reflect the position or policy of the U.S. Department of Veteran Affairs or the Centers for Medicare & Medicaid Services.

**References**


![Figure 4. Final mapping of numeric rating scale (NRS) to verbal descriptor scale (VDS) pain intensity response options. Note: solid lines represent VDS response categories and dotted lines represent NRS response category combinations.](http://biomedgerontology.oxfordjournals.org/)

Whenever possible. However, there is no consensus on the best approach for self-report of pain assessment, and the MDS assessment has not included self-reported measures. As patient and provider preferences for different assessment tools vary, it is useful to provide a choice of scales (20). The results reported in this article use modern measurement theory to provide an empirically based crosswalk between the two most commonly used methods to assess pain intensity, the NRS, and VDS pain scales. The resultant “crosswalk” can be used to compare prevalence rates from the two methods either within or between studies and can also be used to compare results across time points so that both the VDS and the NRS can continue to be used in practice depending on the preference of the clinician and respondent. The similarity of the mapping presented here to the majority of results reported by Jones and colleagues (34) lends credibility to this solution.

The final mapping of the NRS to the VDS was clear for most NRS categories. However, the placement of NRS option 7 in the moderate VDS category was not ideal, as option 7 appeared to straddle the line between VDS moderate and severe categories. Clinicians and researchers may want to keep this in mind as they apply the mapping for various purposes. For example, if the intent is to identify individuals in severe pain, and sensitivity is preferred over specificity, it may be appropriate to classify NRS option 7 as severe. Alternatively, if the goal is to get an optimal estimate of prevalence rates according to the four VDS pain categories, option 7 should be placed within the moderate group. It may also be helpful to consider the impact of pain on functioning when making this classification decision, as recent research has suggested that knowledge of the effect of pain on day-to-day function adds important information, especially for individuals with moderate or severe pain (41).


40. Edelen MO, Reeve B. Applying item response theory (IRT) modeling to questionnaire development, evaluation, and refinement. Qual Life Res. 2007;16(suppl 1):5–18.


