Deborah A. Lowe a  Steve Balsis a  Tyler M. Miller b  Jared F. Benge c  Rachelle S. Doody d

aTexas A & M University, College Station, Tex., bSouth Dakota State University, Brookings, S. Dak., cJack C. Montgomery VA Medical Center, Muskogee, Okla., and dAlzheimer’s Disease and Memory Disorders Center, Department of Neurology, Baylor College of Medicine, Houston, Tex., USA

Abstract

Background/Aims: An item response theory (IRT)-based scoring approach to the Clinical Dementia Rating Scale (CDR) can account for the pattern of scores across the CDR items (domains) and their differential abilities to indicate dementia severity. In doing so, an IRT-based approach can provide greater precision than other CDR scoring algorithms. However, neither a good set of item parameters nor an easily digestible set of instructions needed to implement this approach is readily available. Methods: Participants were 1,326 patients at the Baylor College of Medicine Alzheimer’s Disease and Memory Disorders Clinic. Results: The item parameters necessary for an IRT-based scoring approach were identified (a parameters ranged from 3.01 to 6.22; b parameters ranged from −2.46 to 2.07). Conclusion: This study provides, and demonstrates how to easily apply, IRT-based item parameters for the CDR.

Key Words
Alzheimer’s disease · Assessment · Clinical Dementia Rating Scale · Item response theory · Statistics

Introduction

Determining dementia severity is critical when working with dementia patients, whether providing patient care, establishing the effectiveness of interventions, or conducting translational and clinical research. Depending on the degree of dementia severity of the patients (e.g. mild, moderate or severe), there will be differences in their specific symptom presentations and severities, behavioral and functional outcomes (e.g. aggression, mood disturbance, impairment in activities of daily living), and patterns of cognitive deficits [1–5]. A variety of measures have been developed for determining dementia severity. One of the most widely used measures is the Clinical Dementia Rating Scale (CDR) [6].

The CDR is comprised of six cognitive and functional domains: memory, orientation, judgment and problem solving, community affairs, home and hobbies, and personal care. Clinicians complete the CDR based on interview data from the patient and a collateral source (e.g. spouse, caregiver). According to classic scaling of the measure, the first five domains are scored on a five-point ordinal scale, where 0 = no impairment, 0.5 = question-
able impairment, 1 = mild impairment, 2 = moderate impairment and 3 = severe impairment. The sixth domain (personal care) is scored on a four-point ordinal scale with possible scores of 0, 1, 2 or 3; although clinicians occasionally use a score of 0.5 for this domain, this is outside the realm of traditional scoring practices.

Two primary methods have been used for calculating total scores for the CDR. The most common technique involves a scoring algorithm that weights memory as the primary domain and all other domains as secondary [7]. This algorithm uses the domain scores of a patient to generate a global CDR score and place him or her into one of five stages of dementia severity. Possible global scores are 0, 0.5, 1, 2 and 3 and represent no, questionable, mild, moderate and severe dementia, respectively. Since this approach uses just five stages to represent the full spectrum of dementia severity, it has been critiqued as being a relatively coarse estimation of dementia severity [8]. Researchers have found substantial variability in the degree of dementia severity within particular CDR stages [8–10], indicating that patients within a given CDR stage (e.g. stage 1 = mild dementia) are not equivalent in their underlying degree of dementia severity. In addition, patients with the same degree of dementia severity can sometimes be placed into different CDR stages [8]. It is likely that these problems occur because the scoring algorithm for global CDR scores uses a weighting scheme that is not based on optimal psychometric principles. Consequently, these global scores contain considerable imprecision and only partially capture critical information about true dementia severity.

Another option is to use the sum of the boxes (SOB) scoring approach, in which all domains are weighted equally. The summed score serves as the total estimate of dementia severity, with possible total scores ranging from 0 to 18. Because of its larger range of possible values, the SOB approach is able to detect more subtlety dementia severity than the standard algorithm [10]. Since the SOB approach weights each domain equally, it cannot capture information about the relative ability of each domain to indicate dementia severity. For example, a score of 1 reflecting mild impairment in the memory domain may indicate something different about dementia severity than a score of 1 in the community affairs domain. However, the SOB approach weights both scores for these domains equally. In doing so, the SOB approach, like the traditional scoring algorithm, fails to capture important information about dementia severity.

An item response theory (IRT)-based scoring approach to the CDR has been recommended for increased precision when estimating dementia severity [8, 11]. IRT takes into account the differential ability of each domain to predict and reflect the latent construct (in this case, dementia severity). By accounting for the specific pattern of scores across CDR domains and their differential abilities to indicate dementia severity, more precise estimates of dementia severity can be achieved.

For this approach to gain widespread use, item parameters must be established for the CDR. However, these parameters have yet to be published in the literature. For item parameters to be particularly useful, they should be based on a large sample that is relevant to the population with which the CDR is generally used. The CDR was developed for use with Alzheimer’s disease (AD) and is commonly used in AD clinical trials [7]; it is occasionally used with other forms of dementia and neurocognitive impairment [12]. In addition to using a large, relevant sample, it seems prudent to identify item parameters based on baseline assessments of the patients. This would allow investigators to use those parameters to measure dementia severity when they first assess a patient and to track changes in dementia severity over time. This study fills a critical gap in the existing literature by identifying and providing item parameters for the CDR based on a large sample of patients who were comprehensively assessed for AD and neurocognitive impairment. In addition, we provide a code and easily digestible instructions to enable clinicians and researchers to use these item parameters for increased precision when assessing dementia severity using the CDR.

Methods

Participants
Participants were 1,326 individuals from the Baylor College of Medicine Alzheimer’s Disease and Memory Disorders Clinic (mean age = 73.88 years, SD = 8.72). Education data were missing for 2 participants; for the remaining 1,324 participants the average level of education was 14.37 years (SD = 6.36). Sixty-three percent were female and 91% were Caucasian. Other races represented included African-American (7%) and Asian-American (1%). Diagnoses included probable AD (n = 1,051; 79%), possible AD (n = 187, 14%), and mixed AD (n = 88, 7%). Participants were included for analysis if they had no missing CDR data.

Materials
The CDR was administered to all patients at baseline admission. Six participants received a score of 0 (1%), 497 received a score of 0.5 (37%), 532 received a score of 1 (40%), 221 received a score of 2 (17%) and 70 received a score of 3 (5%).
Procedures
Patients at the Alzheimer's Disease and Memory Disorders Clinic receive comprehensive neurological, neuropsychological and medical testing upon admission (i.e. baseline) and on a yearly basis thereafter. As part of this longitudinal protocol, patients are administered the CDR. Medical diagnoses are based on a consensus conference review of medical records, psychometric test scores and medical evaluations. Additional information on the Alzheimer’s Disease and Memory Disorders Clinic protocol can be found elsewhere [13].

Data Analyses
Data analyses were conducted within an IRT framework. IRT analyses assume unidimensionality of the latent construct (in this case, dementia severity). Although the CDR measures different conceptual constructs (e.g. memory, orientation, personal care), unidimensionality reflects the tendency of those constructs to covary enough that they represent just one statistical factor (i.e. dementia severity) [14, 15]. To test this assumption, we conducted exploratory and confirmatory factor analyses, as described in the Results section below.

We performed the IRT analyses in Multilog v. 6.3 [16], using the graded response model described by Samejima [17, 18],

\[ P(x_j = c|\theta) = \frac{1}{1 + \exp(-a_j(\theta - b_j))} \times \frac{1}{1 + \exp(-a_j(\theta - b_{j+1}))}, \]

where \(a_j\) is the slope, \(b_j\) is threshold for category \(j\), and \(c\) represents response category.

We identified item parameters for each CDR domain. The \(a\) (‘discrimination’) parameter indicates how related that domain is to the latent construct (theta/\(\theta\), i.e. dementia severity). The \(a\) parameter typically ranges from 0 to 3, with larger values indicating sharper discrimination and smaller values indicating weaker discrimination between subjects at any given value of theta [19–21]. The \(b\) (‘difficulty’ or ‘threshold’) parameters indicate the degree of dementia severity at which certain response options have a 50% likelihood of being selected. They also reflect the degree of dementia severity at which domains discriminate most efficiently. Since the CDR domains contain more than two scoring options, there are multiple \(b\) parameters for each domain. For the first five domains (excluding personal care), the \(b1\) parameter indicates the threshold for 50% probability of receiving a score of 0 versus 0.5, 1, 2 or 3. The \(b2\) parameter reflects the threshold for receiving a score of 0 or 0.5 versus 1, 2 or 3, and the \(b3\) parameter indicates the threshold for receiving a score of 0, 0.5 or 1 versus 2 or 3. The \(b4\) parameter reflects the threshold of the latent construct at which there is a 50% probability of receiving a score of 0, 0.5, 1 or 2 versus 3. Since the personal care domain has four scoring options rather than five, its \(b1\) parameter reflects the threshold for 50% probability of receiving a score of 0 versus 1, 2 or 3. Its \(b2\) parameter indicates the threshold for receiving a score of 0 or 1 versus 2 or 3, and its \(b3\) parameter reflects the threshold for receiving a score of 0, 1 or 2 versus 3. Collectively, the \(a\) and \(b\) parameters represent the differential ability of each CDR domain to indicate dementia severity across the full continuum. Using IRT algorithms, these parameters can be used to estimate a person’s dementia severity (i.e. their theta value on the underlying construct of dementia severity). This process can be carried out using the instructions and code we provide in the Conclusions section and Appendix.

Results and Discussion
Results indicated that the data met the assumption of unidimensionality. We conducted an exploratory factor analysis using maximum likelihood estimation SPSS v. 13.0 [22] (see fig. 1 for the scree plot). The results revealed a large first factor (eigenvalue = 4.69, explaining 78% of the variance), with the subsequent factors having eigenvalues substantially less than 1.0. The ratio between the first and second eigenvalues was greater than 10:1. This ratio and the large first factor are highly supportive of the unidimensionality of the CDR. We then performed a confirmatory factor analysis using MPLUS [23] to further test for unidimensionality. According to Hu and Bentler [24], the hypothesized model (in this case, unidimensionality) provides a good fit to the observed data when the Tucker-Lewis index [25] and the comparative fit index [26] are both close to 0.95 (with 1.00 representing a perfect fit). In our confirmatory factor analysis, the Tucker-Lewis index and the comparative fit index were both greater than 0.95 (Tucker-Lewis index + comparative fit index = 0.99), which suggested a good fit between the data and a one-factor model. Given these strong indi-
For the CDR domains (see Table 1). The $a$ parameters ranged from 3.01 to 6.22. These values are notably higher than the typical range of values from 0 to 3, which indicates that the domains are each strongly related to the latent construct (i.e., dementia severity). The $b$ parameters ranged from $-2.46$ to $2.07$.

The $b_1$ parameter reflects the threshold of dementia severity at which each CDR domain has a 50% probability of being scored positively rather than 0. Therefore, this parameter can be thought of as indicating when the abilities measured by each domain become an area of difficulty or deficit. Clinicians may be particularly interested in this parameter because it reflects the differential ability of each CDR domain to detect information about dementia at various levels of the underlying construct of dementia severity. It reflects where along the continuum of dementia severity the clinician is optimally able to detect impairment (whether questionable, mild, moderate or severe) in the abilities assessed by that particular domain (e.g., memory, personal care). Figure 2 shows the $b_1$ parameters for all the CDR domains. Notice that the memory domain has the lowest $b_1$ parameter. This means that it discriminates best at milder degrees of dementia severity; memory problems are likely to be present among milder cases of the disease and to provide the most information about dementia at the mild end of the continuum. On the other hand, note that the personal care curve has the highest $b_1$ parameter, which indicates that difficulties in personal care provide discriminative information about dementia at the more severe end of the spectrum. Also notice that the remaining curves are grouped together. This means that their abilities to discriminate occur relatively simultaneously and between memory and personal care along the dementia severity spectrum.

CI = Confidence interval; JPS = judgment and problem solving; CA = community affairs; HH = home and hobbies; PC = personal care.

**Table 1.** Item parameters for the CDR

<table>
<thead>
<tr>
<th>Domain</th>
<th>$a$</th>
<th>95% CI</th>
<th>$b_1$</th>
<th>95% CI</th>
<th>$b_2$</th>
<th>95% CI</th>
<th>$b_3$</th>
<th>95% CI</th>
<th>$b_4$</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Memory</td>
<td>3.88</td>
<td>3.47, 4.29</td>
<td>-2.46</td>
<td>-2.71, -2.21</td>
<td>-1.11</td>
<td>-1.19, -1.03</td>
<td>0.46</td>
<td>0.38, 0.54</td>
<td>1.71</td>
<td>1.59, 1.83</td>
</tr>
<tr>
<td>Orientation</td>
<td>3.55</td>
<td>3.22, 3.88</td>
<td>-1.41</td>
<td>-1.51, -1.31</td>
<td>-0.66</td>
<td>-0.72, -0.60</td>
<td>0.52</td>
<td>0.44, 0.60</td>
<td>1.69</td>
<td>1.57, 1.81</td>
</tr>
<tr>
<td>JPS</td>
<td>4.71</td>
<td>4.26, 5.16</td>
<td>-1.44</td>
<td>-1.52, -1.36</td>
<td>-0.56</td>
<td>-0.62, -0.50</td>
<td>0.74</td>
<td>0.66, 0.82</td>
<td>1.54</td>
<td>1.44, 1.64</td>
</tr>
<tr>
<td>CA</td>
<td>6.22</td>
<td>5.63, 6.81</td>
<td>-1.22</td>
<td>-1.28, -1.16</td>
<td>-0.40</td>
<td>-0.46, -0.34</td>
<td>0.72</td>
<td>0.66, 0.78</td>
<td>2.05</td>
<td>1.93, 2.17</td>
</tr>
<tr>
<td>HH</td>
<td>4.88</td>
<td>4.41, 5.35</td>
<td>-1.22</td>
<td>-1.30, -1.14</td>
<td>-0.42</td>
<td>-0.48, -0.36</td>
<td>0.71</td>
<td>0.63, 0.79</td>
<td>1.69</td>
<td>1.59, 1.79</td>
</tr>
<tr>
<td>PC</td>
<td>3.01</td>
<td>2.68, 3.34</td>
<td>0.21</td>
<td>0.13, 0.29</td>
<td>1.05</td>
<td>0.95, 1.15</td>
<td>2.07</td>
<td>1.89, 2.25</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Fig. 2.** Differential abilities of CDR domains to discriminate dementia severity ($b_1$ parameters). JPS = Judgment and problem solving; CA = community affairs; HH = home and hobbies; PC = personal care.
Clinicians and researchers can use the item parameters we identified to estimate the dementia severity of patients using the CDR. We provide the required code and simple instructions for using these item parameters in the Conclusions section and Appendix.

**Conclusions**

This study identifies and provides item parameters for the CDR that can be used for greater precision when estimating dementia severity. Exploratory and confirmatory factor analyses supported the unidimensionality of the CDR. Although the CDR measures different conceptual constructs (i.e., memory, orientation, judgment, and problem solving, community affairs, home and hobbies, personal care), these domains covary enough that they represent one statistical factor (i.e., dementia severity). IRT analyses indicated that all six domains are highly related to the latent construct of dementia severity (as seen by the a parameters). By examining the b1 parameters, it is evident that the memory domain best discriminates at milder levels and personal care differentiates at more severe levels of dementia severity, with the other four domains discriminating at intermediate degrees of dementia severity. This finding converges with what is already known about the process of deterioration in AD, namely, that memory deficits tend to be present relatively early in the course of the disease, and impairment in activities of daily living, such as the ability to care for oneself, is evident at more severe degrees of dementia [27, 28]. An IRT-based scoring of the CDR provides rich information about a patient's precise degree of dementia severity by accounting for his or her pattern of scores and the differential abilities of the domains to optimally discriminate at various levels of dementia severity. The set of item parameters we provide in this paper can guide clinicians in how to use the CDR to more precisely measure dementia severity.

Once item parameters have been provided for a test, investigators can then use those parameters to interpret test scores within a similar sample. It is important to note that our sample was comprised of outpatient adults presenting with cognitive concerns to a memory disorders clinic, who were subsequently diagnosed with probable, possible or mixed AD. In addition, our sample was highly educated, predominantly Caucasian and generally mildly to moderately impaired (94% were given CDR scores of 0.5, 1 or 2). The educational and racial make-up of this sample limits its generalizability to demographically diverse samples. As more data are collected across demographic dimensions (gender, race, etc.), investigators will need to explore responsibly the influence of these variables on item parameter estimation.

The scarcity of CDR scores of 0 or 3 in our sample was consistent with our intentions. Among patients who are initially presenting for treatment related to cognitive concerns, scores of 0 (i.e. no dementia) or 3 (i.e. severe dementia) are rare. Thus, the impairment levels in this sample are appropriately generalizable to other groups of patients presenting at baseline for cognitive concerns. This study, therefore, enables clinicians and researchers to utilize an IRT-based scoring method for the CDR relative to our sample of patients with AD presenting with cognitive concerns to a memory disorders clinic.

We have provided all the necessary information in the Appendix to do this, using Multilog software [16]. There are four parts to the Appendix: (1) the code for data analysis (IRTSCORE.MLG), (2) the parameter file (IRTSCORE.PAR), (3) the sample data file (IRTSCORE.TXT) and (4) instructions for using this IRT-based scoring approach. Using the instructions we provide, the process of using an IRT-based scoring approach is quite simple.

By following the steps we provide in the Appendix, investigators can easily obtain theta scores for individuals that indicate their precise degree of dementia severity. Theta scores are a standardized index of dementia severity with a mean of 0 and SD of 1. Negative theta values represent milder degrees of dementia severity and positive values indicate more severe degrees of dementia. Because theta is standardized, different values of theta can be interpreted in relation to each other. This means that a patient’s dementia severity at different time points (e.g. longitudinal assessment) can be compared and that estimates of dementia severity for multiple patients can also be compared. For example, a theta value of 1.79 represents dementia that is more severe than a theta value of 0.52 (regardless of whether these theta values represent data for 1 or 2 individuals). Similarly, a theta value of 0.52 indicates dementia that is more severe than a theta value of –0.97. Thus, using the item parameters, code and instructions we provide allows for the comparison of dementia severity between different patients and across multiple time points.

IRT adds much-needed precision to the measurement of dementia severity using the CDR. Scores from tests like the CDR always include some degree of measurement error in addition to the patient’s ‘true score’ (e.g. a patient’s CDR score consists of his or her true dementia severity...
Item Parameters for the CDR

Dement Geriatr Cogn Disord 2012;34:128–134

Appendix

Instructions for an IRT-Based Scoring Approach Using Multilog Software

(1) Save the three documents located below together in a folder according to the file names provided (i.e. IRTSCORE.MLG, IRTSCORE.PAR and IRTSCORE.TXT).

(2) Insert the CDR data of the patients into the IRTSCORE.TXT file using the provided format (i.e. domain scores; see note below about recoding).

(3) Open the IRTSCORE.MLG file in Multilog.

(4) Run the .MLG code (File/Run), which then provides theta values (i.e. dementia severity) for each case entered.

This software program only accepts integer codes. Because the first five CDR domains contain a 0.5 code, the codes for those five domains will need to be recoded as follows: 0 remains 0, 0.5 becomes 1, 1 becomes 2, 2 becomes 3, and 3 becomes 4. The codes for the personal care domain do not need to be recoded, since they are already integers. For example, consider a person whose pattern of scores across CDR domains is 0.5/0/1/0.5/0/0. After recoding the appropriate domains, the scores become 1/0/2/1/0/0. Using the code and instructions we provide, entering this pattern of domain scores in Multilog returns a theta value of –1.12. Recoding is a simple and quick process that allows the program to function as intended.

Code for data analysis (IRTSCORE.MLG):
MULTILOG for Windows
Created on: XX June 2012, 14:23:53
>PROBLEM SCORE, INDIVIDUAL, NITEMS = 6, NGROUPS = 1, NEXAMINEES = 1, DATA = 'IRTSCORE.TXT';
>TEST ALL, GRADED, NC=(5,5,5,5,5,4);
>START ALL, PARMS = 'IRTSCORE.PAR';
>END;

Parameter file for data analysis (IRTSCORE.PAR):
-3.88256 -2.45973 -1.11295 0.46317 1.71387
-3.55073 -1.40703 -0.66212 0.51609 1.69051
-4.70577 -1.44390 -0.55783 0.74408 1.54372
-6.21514 -1.22130 -0.40108 0.72334 2.04728
-4.88379 -1.22151 -0.41967 0.70814 1.69043
-3.01493 -0.20990 1.05454 2.07345

Data file for data analysis (IRTSCORE.TXT):
102100

Final Note
Multilog allows the user to specify a prior distribution for the item parameters, but this is often not done in practice [Edwards, pers. commun., 7 June, 2012]. In fact, with a large enough sample size, the choice of a prior distribution often does not have a noteworthy effect on the resulting item parameter estimates.

Acknowledgements

This study was supported by the Cynthia and George Mitchell Foundation (S.B. and J.F.B.), the Cain Foundation (R.S.D.) and NIA Grant 1R03AG039663–01A1 (S.B.).
References


