Item Response Analysis of Alzheimer's Disease Assessment Scale

Nishant Verma and Mia K. Markey

Abstract— **Alzheimer's Disease Assessment Scale-cognitive subscale (ADAS-cog) suffers from low sensitivity in detecting changes in Alzheimer's disease progression in clinical trials of disease-modifying treatments. A comprehensive psychometric analysis of the items in ADAS-cog assessment can help in identifying and improving the insensitive items. Item response theory provides a suitable framework for investigating the ADAS-cog items; however, it requires prior information on the underlying latent construct for reliable analysis. In this study, we perform an exploratory item response analysis to investigate the latent construct underlying the relationships between the ADAS-cog item responses. The results indicate that the underlying latent construct of ADAS-cog is multidimensional in nature with the latent factors measuring cognitive declines in several domains (such as memory, praxis, and language domains).**

I. INTRODUCTION

The Alzheimer's Disease Assessment Scale-cognitive subscale (ADAS-cog) is the standard assessment tool for measuring Alzheimer's progression in clinical trials of disease-modifying treatments [1]. However, ADAS-cog has been reported to be highly insensitive to changes in disease progression in clinical trials [2]. This is a major reason behind the failure of all clinical trials to date of disease-modifying treatments of Alzheimer's disease. The ADAS-cog assessment consists of 11 items measuring various cognitive abilities (such as memory and language), which are considered to be early hallmarks of Alzheimer's disease. Patients are scored on each individual item, which are added to obtain the total ADAS-cog scores. The possible scores on ADAS-cog assessment range from 0 to 70, where higher scores signify greater cognitive impairment. Previous studies have indicated that individual ADAS-cog items may have varying abilities and limitations in detecting cognitive impairment. Therefore, a thorough psychometric analysis of ADAS-cog is essential for its optimal application in clinical trials.

Item response theory (IRT) is a class of latent variable models, which provides a suitable framework for a thorough psychometric analysis of ADAS-cog assessment items. IRT links categorical manifest variables (such as ADAS-cog item responses) to latent factors (such as Alzheimer's disease severity), which can not be directly measured. Some previous studies have explored the application of IRT in analyzing ADAS-cog assessment items [3]. However, these studies suffered from several limitations, primarily due to decisions made during the exploratory phase of IRT modeling.

The estimation of IRT models is based on a set of fairly strong assumptions: (i) local independence between the item responses given a subject's latent trait value, (ii) nature of the item characteristic curves, and (iii) parameter invariance. If these assumptions are not reasonably met, the validity of IRT estimates (especially latent factors) becomes severely compromised. All previous studies considered the underlying latent structure in ADAS-cog assessment to be unidimensional, i.e., they assumed that a single latent factor explains the relationships between all ADAS-cog items. Prior studies concluded that the ADAS-cog exhibited unidimensionality based on a set of techniques (such as principal component analysis and parallel analysis), which had been reported previously to suffer from inconsistency in determining the number of latent factors for common factor analysis [4]. The failure to account for multidimensionality in the latent construct of ADAS-cog assessment can potentially translate to local dependence between the item responses, rendering the IRT analysis invalid. The objective of this study is to investigate the underlying latent construct of the ADAS-cog assessment. This basically involves establishing the required number of latent factors, interpreting them and their loadings on the ADAS-cog item responses.

II. METHODS

A. Item Response Theory

IRT probabilistically models a subject's responses to an item by specifying the manner in which the subject's latent factors interact with the characteristics of that item. For a dichotomous item with 2 possible response categories, IRT models the probability of answering it correctly as

$$P(x_{ij}=1|\boldsymbol{\theta_i},\boldsymbol{\alpha_j},d_j)=\frac{1}{1+\exp[-(\boldsymbol{\alpha_j^T}\boldsymbol{\theta_i}+d_j)]} \quad (1)$$

where, $i=1,\dots,N$ represent the distinct participants and $j=1,\dots,n$ represent the test items in the instrument. $\boldsymbol{\theta_i}=(\theta_{i1},\dots,\theta_{im})$ are the m latent factors with associated item slopes $\boldsymbol{\alpha_j}=(\alpha_1,\dots,\alpha_m)$ and d_j is the item intercept. For a polytomous item j with $C_j\geq 2$ response categories $k=\{0,\dots,C_j-1\}$, the boundaries of response probabilities are

$$P(x_{ij}\geq 0|\boldsymbol{\theta_i},\boldsymbol{\alpha_j},\boldsymbol{d_j})=1,$$
$$P(x_{ij}\geq 1|\boldsymbol{\theta_i},\boldsymbol{\alpha_j},\boldsymbol{d_j})=\frac{1}{1+\exp[-(\boldsymbol{\alpha_j^T}\boldsymbol{\theta_i}+d_1)]},$$

N. Verma is with the Dept. of Biomedical Engineering, The University of Texas at Austin, Austin, Texas, 78712, USA and NeuroTexas Institute, St. David's HealthCare, Austin, TX, 78705. vnishant@utexas.edu

M. K. Markey is with the Dept. of Biomedical Engineering, The University of Texas at Austin, Austin, Texas, 78712 and Dept. of Imaging Physics, The University of Texas MD Anderson Cancer Center, Houston, Texas, 77030. mia.markey@utexas.edu

$$P(x_{ij} \geq 2|\boldsymbol{\theta_i}, \boldsymbol{\alpha_j}, \boldsymbol{d_j}) = \frac{1}{1 + \exp[-(\boldsymbol{\alpha_j^T}\boldsymbol{\theta_i} + d_2)]},$$
$$\vdots$$
$$P(x_{ij} \geq C_j|\boldsymbol{\theta_i}, \boldsymbol{\alpha_j}, \boldsymbol{d_j}) = 0$$

where, $\boldsymbol{d_j} = (d_1, \ldots, d_{(C_j-1)})$ are the intercepts corresponding to the boundaries of response probabilities. These boundaries can be used to obtain the conditional probability for any item response $x_{ij} = k$ as

$$P(x_{ij} = k|\boldsymbol{\theta_i}, \boldsymbol{\alpha_j}, \boldsymbol{d_j}) = P(x_{ij} \geq k|\boldsymbol{\theta_i}, \boldsymbol{\alpha_j}, \boldsymbol{d_j}) - P(x_{ij} \geq k+1|\boldsymbol{\theta_i}, \boldsymbol{\alpha_j}, \boldsymbol{d_j})$$

B. Model Estimation

Expressing the response data x_{ij} in an indicator form χ, where $\chi(x_{ij}, k) = \begin{cases} 1, & \text{if } x_{ij} = k \\ 0, & \text{otherwise} \end{cases}$, and defining $\boldsymbol{\Psi}$ as the set of all unknown item parameters, the conditional likelihood for the i^{th} response vector $\boldsymbol{x_i}$ can be defined as

$$L(\boldsymbol{x_i}|\boldsymbol{\Psi}, \boldsymbol{\theta}) = \prod_{j=1}^{n} \prod_{k=0}^{C_j-1} P(x_{ij} = k|\boldsymbol{\Psi}, \boldsymbol{\theta_i})^{\chi(x_{ij},k)}$$

IRT assumes a multivariate normal distribution $g(\theta)$ over the latent variables $\boldsymbol{\theta}$ and integrates them out of the likelihood function. Therefore, the marginal likelihood function of the observed data $\boldsymbol{X} = (\boldsymbol{x_1}, \ldots, \boldsymbol{x_N})$ becomes

$$L(\boldsymbol{\Psi}|\boldsymbol{X}) = \prod_{i=1}^{N} \left[\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} L(\boldsymbol{x_i}|\boldsymbol{\Psi}, \boldsymbol{\theta}) g(\boldsymbol{\theta}) d\boldsymbol{\theta} \right]$$

The recommended method for the estimation of IRT models is the expectation maximization (EM) algorithm using fixed Gauss-Hermite quadrature.

III. Experiments & Results

A. Data

The data were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (`http://adni.loni.ucla.edu/`). The data consisted of item-level ADAS-cog assessment response data from 671 mild cognitively impaired (MCI) patients and 870 Alzheimer's disease (AD) patients (including MCI patients that converted to AD). While longitudinal data are available on these patients, only a single time point data was used from each patient in this study in order to minimize individual level effects, i.e., minimize correlated responses to the items in ADAS-cog assessment. The specific time point for each patient was chosen at random to obtain a uniform distribution of ADAS-cog assessment data across all disease severity levels. Besides the 11 ADAS-cog items, we also include delayed word recall, which has been shown to be important for measuring memory impairment early in the course of dementia. Therefore, the data contained response data from 1275 patients on 12 ADAS-cog items: word recall task (Q1), commands (Q2), constructional praxis (Q3), delayed word recall (Q4), naming objects and fingers (Q5), ideational praxis (Q6), orientation (Q7), word recognition task (Q8), remembering test instructions (Q9), comprehension of spoken language (Q10), word finding difficulty (Q11), and spoken language (Q12).

B. Exploratory Item Response Analysis

We perform an exploratory item response analysis to understand the underlying latent construct of ADAS-cog items i.e., the number of cognitive domains being measured by ADAS-cog. The selection of an appropriate number of latent factors is an extremely important decision in IRT modeling. Since the response data for all ADAS-cog items are ordinal in nature, we calculate pairwise polychoric correlations between the items. All item pairs show significant correlations (> 0.35) except the pairs of Q2 & Q3, Q3 & Q8, Q3 & Q5 and Q3 & Q11. In general, item Q3 illustrates low pairwise correlations with other items (Fig. 1). Several techniques are commonly used for determining the number of latent factors in an exploratory factor analysis:

- Kaiser's Rule: The most common practice has been to consider the components with eigenvalues greater than 1. The eigenvalues of the polychoric correlation matrix are 6.30, 1.29, .88, .69, .57, .52, .42, .35, .33, .26, .23 and .16. Kaiser's rule suggests that $m = 2$ latent factors are sufficient since only two eigenvalues ≥ 1.
- Scree Plot: Visual inspection of the scree plot can be used to determine the appropriate number of factors based on sharp breaks in the plot. This method also suggests considering $m = 2$ latent factors.
- Parallel Analysis: Parallel analysis determines the number of factors by comparing the scree of factors of the observed data with that of random data. For ADAS-cog items, parallel analysis suggests $m = 4$ as the appropriate number of latent factors.

As frequently reported in literature, different techniques suggest different number of latent factors for common factor models [4]. Therefore, as recommended by [4], we investigate four solutions (S1, S2, S3, and S4) with the number of latent factors as $m = \{1, 2, 3, 4\}$. We compare between these solutions to select the most meaningful latent construct using the following criteria: (a) assessment of model fit, (b) the

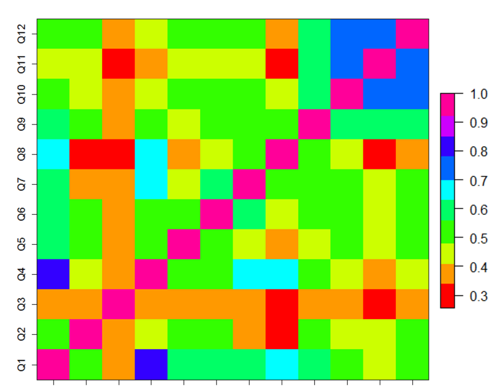

Fig. 1: Polychoric correlations between ADAS-cog items.

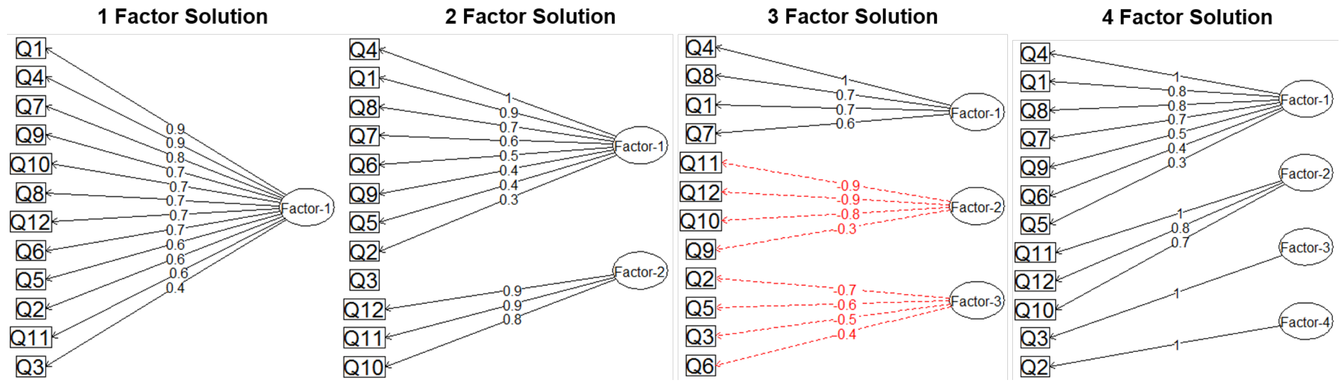

Fig. 2: Latent factor loadings in 1 factor model (S1), 2 factor model (S2), 3 factor model (S3), and 4 factor model (S4).

TABLE I: Comparing model fit statistics between S1, S2, S3, and S4. The statistics considered are Log-Likelihood (Lok-Lik), Akaike Information Criterion (AIC), sample size corrected AIC (AICc), Bayesian Information Criterion (BIC) and Sample-Size Adjusted BIC (SABIC) .

Statistic	S1	S2	S3	S4
Log-Lik	-15361.7	-15137.9	-15098.6	-14944.9
AIC	30871.5	30445.7	30387.2	30097.8
AICc	30880.7	30458.0	30402.7	30116.5
BIC	31252.6	30883.5	30876.5	30633.5
SABIC	31017.6	30613.5	30574.8	30303.1

	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12
Q1						S3,S4					S1, S2, S3, S4	
Q2			S1, S2, S3, S4		S1, S2, S3, S4	S1, S2, S3, S4				S1, S2, S3, S4	S1, S2, S3, S4	S1, S2, S3, S4
Q3						S1, S2, S3, S4	S1, S2, S3, S4			S4	S1, S2, S3, S4	
Q4												
Q5						S1, S2, S3, S4					S1	
Q6												
Q7											S1, S2, S3, S4	
Q8												
Q9										S1	S1, S4	S1
Q10											S1	S1
Q11												S1
Q12												

Fig. 3: Matrix showing the item pairs where S1, S2, S3, and S4 rejected the null hypothesis of local independence.

validity of local independence assumption, (c) the validity of parameter invariance assumption, (d) the interpretation of estimated latent factors, and (e) the structure of latent factor loadings on the ADAS-cog items. The assumption of local independence is tested using the G^2 statistic [5], which has been demonstrated to be highly sensitive in detecting multidimensionality.

First, we assess the model fit for the four solutions S1, S2, S3, and S4 (Table I) and perform model comparisons between the solutions using χ^2 statistics based on the Log-Likelihood statistics. In comparison with S1, S2 shows significantly better model fit (p-value = 0, $\chi^2 = 447.74$ with degrees of freedom $df = 11$). When S3 is compared with S2, S3 has significantly better model fit (p-value=0, $\chi^2 = 78.52$ with $df = 10$). When S4 is compared with S3, S4 shows significantly better fit (p-value = 0, $\chi^2 = 307.39$ with $df = 9$). While increasing the dimensionality is producing better fitting models, model selection solely based on the model fit statistics is misleading. The selection of appropriate model complexity should be performed based on a combination of criteria discussed above.

Next, we evaluate the structure and interpretations of the factor loadings in the four solutions. Fig. 2 shows the factor loadings on ADAS-cog items for solutions S1, S2, S3, and S4. The unidimensional solution (S1) has a simple structure and the latent factor can be interpreted as an indicator of patient's decline across all cognitive domains. However, S1 rejects the local independence hypothesis between several item pairs (Fig. 3). Specifically, the language-related items (Q9, Q10, Q11, Q12) illustrate local dependence with each other, indicating the presence of additional latent factors.

The inclusion of a second latent factor (S2) eliminates the local dependence between the language-related items (Q9, Q10, Q11, and Q12). The second latent factor also shows heavy loadings on most of these language-related items (Fig. 2) and, therefore, can be interpreted as an indicator of decline in patient's language abilities. However, S2 still rejects the null hypothesis of local independence for several item pairs (Fig. 3). In S2, items Q5, Q6, Q3, and Q2 have factor loadings of less than 0.5. The introduction of a 3^{rd} latent factor (S3) causes these 4 items to form a separate item cluster, dominantly explained by the third latent factor.

An inspection of the factor loadings (Fig. 2) in S3 reveals an interesting structure among the ADAS-cog items: 1^{st} factor loads dominantly on all memory-related items (Q1, Q4, Q7, Q8), 2^{nd} factor loads dominantly on all language-related items (Q10, Q11, Q12), and the 3^{rd} factor loads dominantly on all praxis-related items (Q2, Q3, Q6). Therefore, the 3 factors are interpreted as indicators of a patient's decline in the memory, language and praxis domains, respectively.

Items Q9 (remembering test instructions) is not dominantly explained by any single latent factor. Q9 illustrates loadings of 0.33 from the 1^{st} factor and 0.34 from the 2^{nd} factor, which makes sense since Q9 involves both understanding the test instructions (language-component)

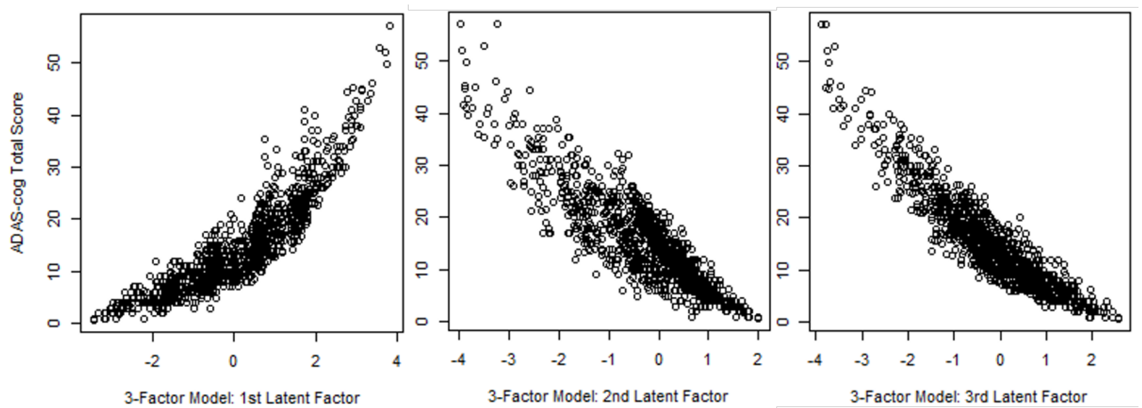

Fig. 4: Scatterplots of the latent factors estimated in the three factor solution (S3) against the total ADAS-cog scores.

and remembering them (memory-component). Q5 (naming objects and fingers), while being related to memory, shows dominant loading of 0.52 from the 3^{rd} factor. This is unusual since all other items in that domain are praxis-related. This may indicate that Q5 measures some latent factor that is not measured by other ADAS-cog items.

Both S3 and S4 do not reduce the local dependence between item responses (in Fig. 3) and reject the local independence assumption for all the same item pairs as S2. Furthermore, S4 shows new items pairs (Q1 & Q6, Q3 & Q10, and Q9 & Q11) with local dependence, indicating an overestimation of the number of latent factors. The overestimation can also be concluded from the factor loadings of S4 (in Fig. 2), where the 3^{rd} and 4^{th} factors load primarily only on the Q3 and Q2 items. This indicates overestimation since the latent factors are modeling individual items rather than modeling the common trait underlying several items.

Based on the criteria discussed above, both the two factor (S2) and the three factor (S3) solutions seem appropriate. For better interpretation, we have included the scatterplots of the estimated latent factors in S2 (in Fig. 5) and S3 (in Fig. 4) against the total ADAS-cog scores. The latent factors corresponding to both S2 and S3 solutions show smoothly changing nonlinear profiles against the total ADAS-cog scores. This is very promising since it indicates that proper modeling of response data from ADAS-cog assessment can potentially improve its sensitivity in clinical trials of disease-modifying treatments. As observed in Fig. 2, the latent factors of S3 correspond to indicators of cognitive decline in memory (1st factor), language (2nd factor), and praxis (3rd factor) domains, respectively. S2 on the other hand, combines memory and praxis domains into the same latent factor (1st factor of S2). The 2nd latent factor of S2 represents the decline in language domain, which is same as the 2nd factor of S3. This can also be concluded from their scatterplots, which appear very similar except the direction of change against the ADAS-cog scores.

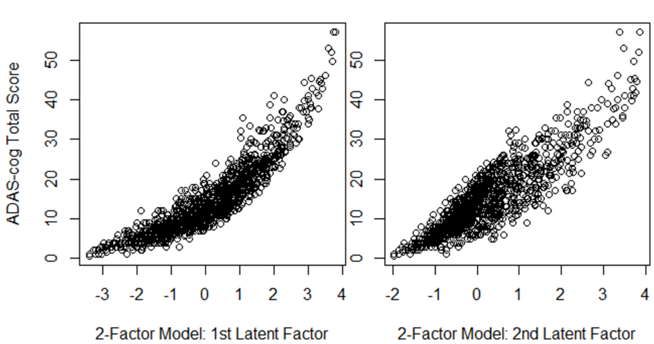

Fig. 5: Scatterplots of the latent factors estimated in the two factor solution (S2) against the total ADAS-cog scores.

IV. CONCLUSIONS

This study explored the underlying latent construct of ADAS-cog assessment as part of the exploratory item response analysis. The results suggest that the latent construct is not unidimensional and requires more than one latent factor for modeling ADAS-cog response data. However, even after inclusion of multiple latent factors, local dependencies are observed between several item pairs. Therefore, as future work, we will investigate the effect of local dependencies on parameter estimation of IRT models. We will use the latent construct developed in this study for also analyzing individual items' abilities in measuring Alzheimer's disease progression.

REFERENCES

[1] W. G. Rosen, R. C. Mohs, and K. L. Davis, "A new rating scale for Alzheimer's disease." *The American journal of psychiatry*, 1984.

[2] N. Raghavan, M. N. Samtani, M. Farnum, E. Yang, G. Novak, M. Grundman, V. Narayan, and A. DiBernardo, "The adas-cog revisited: Novel composite scales based on adas-cog to improve efficiency in mci and early ad trials," *Alzheimer's & Dementia*, vol. 9, no. 1, pp. S21–S31, 2013.

[3] S. Ueckert, E. L. Plan, K. Ito, M. Karlsson, B. Corrigan, and A. C. Hooker, "Improved Utilization of ADAS-cog Assessment Data through Item Response Theory based Pharmacometric Modeling," *Pharmaceutical research*.

[4] J. M. Conway and A. I. Huffcutt, "A review and evaluation of exploratory factor analysis practices in organizational research," *Organizational research methods*, vol. 6, no. 2, pp. 147–168, 2003.

[5] W.-H. Chen and D. Thissen, "Local dependence indexes for item pairs using item response theory," *Journal of Educational and Behavioral Statistics*, vol. 22, no. 3, pp. 265–289, 1997.