

## RESEARCH ARTICLE

# State, trait, and accumulated features of the Alzheimer's Disease Assessment Scale Cognitive Subscale (ADAS-Cog) in mild Alzheimer's disease

Hugo Cogo-Moreira<sup>1</sup> | Saffire H. Krance<sup>2,3,4</sup> | Che-Yuan Wu<sup>3,4,5</sup> |  
Krista L. Lanctôt<sup>3,5,6,7,8</sup> | Nathan Herrmann<sup>4,6</sup> | Sandra E. Black<sup>3,4,9</sup> |  
Bradley J. MacIntosh<sup>3,4,10</sup> | Jennifer S. Rabin<sup>4,11,12</sup> | Michael Eid<sup>13</sup> |  
Walter Swardfager<sup>3,4,5,8</sup> | for the Alzheimer's Disease Neuroimaging Initiative<sup>1</sup>

<sup>1</sup>Department of Education, ICT and Learning, Østfold University College, Halden, Norway

<sup>2</sup>Schulich School of Medicine and Dentistry, University of Western Ontario, London, Ontario, Canada

<sup>3</sup>Sandra Black Centre for Brain Resilience and Recovery, Sunnybrook Research Institute, Toronto, Ontario, Canada

<sup>4</sup>Hurvitz Brain Sciences Program, Sunnybrook Research Institute, Toronto, Ontario, Canada

<sup>5</sup>Department of Pharmacology & Toxicology, University of Toronto, Toronto, Ontario, Canada

<sup>6</sup>Department of Psychiatry, University of Toronto, Toronto, Ontario, Canada

<sup>7</sup>Division of Neurology, Department of Medicine, Sunnybrook Health Sciences Centre, Toronto, Ontario, Canada

<sup>8</sup>KITE Toronto Rehabilitation Institute, University Health Network, Toronto, Ontario, Canada

<sup>9</sup>Department of Neurology, University of Toronto, Toronto, Ontario, Canada

<sup>10</sup>Department of Medical Biophysics, University of Toronto, Toronto, Ontario, Canada

<sup>11</sup>Harquail Centre for Neuromodulation, Sunnybrook Research Institute, Toronto, Ontario, Canada

<sup>12</sup>Rehabilitation Sciences Institute, University of Toronto, Toronto, Ontario, Canada

<sup>13</sup>Department of Education and Psychology, Freie Universität Berlin, Berlin, Germany

## Correspondence

Walter Swardfager, 1 King's College Circle,  
Toronto, ON M5S 1A8, Canada.

Email: [w.swardfager@utoronto.ca](mailto:w.swardfager@utoronto.ca)

Alzheimer's Disease Neuroimaging Initiative data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database ([adni.loni.usc.edu](http://adni.loni.usc.edu)). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: [http://adni.loni.usc.edu/wp-content/uploads/how\\_to\\_apply/ADNI\\_Acknowledgement\\_List.pdf](http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf)

## Abstract

**Background:** The Alzheimer's Disease Assessment Scale Cognitive Subscale (ADAS-Cog) is used to assess decline in memory, language, and praxis in Alzheimer's disease (AD).

**Methods:** A latent state–trait model with autoregressive effects was used to determine how much of the ADAS-Cog item measurement was reliable, and of that, how much of the information was occasion specific (state) versus consistent (trait or accumulated from one visit to the next).

**Results:** Participants with mild AD ( $n = 341$ ) were assessed four times over 24 months. Praxis items were generally unreliable as were some memory items. Language items were generally the most reliable, and this increased over time. Only two ADAS-Cog

Hugo Cogo-Moreira and Saffire H. Krance contributed equally to this study.

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs](https://creativecommons.org/licenses/by-nc-nd/4.0/) License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2023 The Authors. *Alzheimer's & Dementia: Translational Research & Clinical Interventions* published by Wiley Periodicals LLC on behalf of Alzheimer's Association.

items showed reliability  $>0.70$  at all four assessments, word recall (memory) and naming (language). Of the reliable information, language items exhibited greater consistency (63.4% to 88.2%) than occasion specificity, and of the consistent information, language items tended to reflect effects of AD progression that accumulated from one visit to the next (35.5% to 45.3%). In contrast, reliable information from praxis items tended to come from trait information. The reliable information in the memory items reflected more consistent than occasion-specific information, but they varied between items in the relative amounts of trait versus accumulated effects.

**Conclusions:** Although the ADAS-Cog was designed to track cognitive decline, most items were unreliable, and each item captured different amounts of information related to occasion-specific, trait, and accumulated effects of AD over time. These latent properties complicate the interpretation of trends seen in ordinary statistical analyses of trials and other clinical studies with repeated ADAS-Cog item measures.

#### KEYWORDS

Alzheimer's Disease Assessment Scale Cognitive Subscale, Alzheimer's disease, cognition, latent state–trait autoregressive model, structural equation modelling

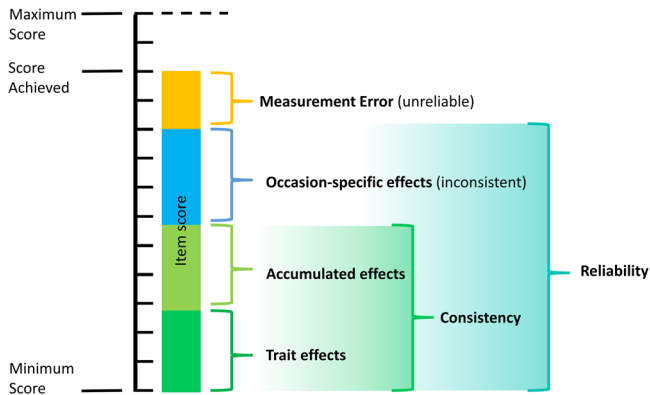
#### Highlights

- Studies have described unfavorable psychometric properties of the Alzheimer's Disease Assessment Scale Cognitive Subscale (ADAS-Cog), bringing into question its ability to track changes in cognition uniformly over time. There remains a need to estimate how much of the ADAS-Cog measurement is reliable, of that how much is occasion specific versus consistent, and of the consistent information, how much represents enduring traits versus autoregressive effects (i.e., effects of Alzheimer's disease [AD] progression carried over from one assessment to the next).
- A latent state–trait model with autoregressive effects in mild AD found most items to be unreliable, and each item to capture different amounts of occasion-specific, trait, and autoregressive information. Language items, specifically, naming and the memory item word recall, were the most reliable.
- Psychometric idiosyncrasies of individual items complicate the interpretation of their summed score, biasing ordinary statistical analyses of repeated measures in mild AD. Future studies should consider item trajectories individually.

## 1 | INTRODUCTION

The Alzheimer's Disease Assessment Scale Cognitive Subscale (ADAS-Cog), designed to assess the severity of cognitive dysfunction in Alzheimer's disease (AD),<sup>1</sup> has been a mainstay in AD studies since its creation in 1984.<sup>2</sup> Given its widespread use, many studies have aimed to evaluate,<sup>1,3,4</sup> modify,<sup>5,6</sup> and optimize<sup>7–9</sup> the ADAS-Cog for its various applications. The commonest application of the ADAS-Cog has been to track AD progression over time in clinical trials and observational studies.

Several studies have questioned the ability of the ADAS-Cog to track changes over time reliably.<sup>10</sup> Discrepancies have been described between changes on the ADAS-Cog scores and clinical improvement on measures such as the Clinician's Interview-Based Impression of Change Plus caregiver input and the Goal Attainment Scaling.<sup>11</sup> Studies have also described some unfavorable psychometric characteristics, including ceiling and floor effects,<sup>7,12,13</sup> and poor test–retest reliability among 7 of 11 items including “following commands” (intraclass correlation coefficient [ICC] = 0.44), “ideational praxis” (ICC = 0.58), “word recognition” (ICC = 0.60), “spoken language ability” (ICC = 0.68),



**FIGURE 1** Visual depiction of the decomposition of a hypothetical assessment score provided by the latent state-trait autoregressive model.

“word-finding difficulty” (ICC = 0.69), and “comprehension” (ICC = 0.66).<sup>14</sup> The minimum standards for reliability for research purposes are often considered to be ICC values >0.6 to 0.8.<sup>15–17</sup> However, to guide clinical decision making, values of 0.90<sup>18</sup> or 0.95<sup>19</sup> would be considered cut-offs.

In a mild AD population, a recent study evaluated the ADAS-Cog for longitudinal invariance, a statistical criterion required before means can be compared from one moment of assessment to the next.<sup>20</sup> Those results raised concerns that the different items might track AD progression differently over time. Although typically the ADAS-Cog items are summed to arrive at total or subdomain scores, that analysis also suggested that this might be discouraged because the items did not inform an underlying cognitive factor, or underlying subdomain factors (e.g., memory, praxis, language) uniformly from one visit to the next. Nonetheless, that result does not preclude the use of individual ADAS-Cog items to track the progression of AD over time or to compare their trajectories to other features of AD. It may be that different symptoms of AD progress differently. Llano et al.<sup>7</sup> noted that some items differentiated healthy controls, people with mild cognitive impairment (MCI), and people with AD very well, while others did not, suggesting that the items may have different properties to capture AD progression at different stages of disease. To distinguish between these possibilities, the reliability of the items needs to be assessed in the context of their longitudinal trajectories. In contrast to previous work, the aim of this study is therefore to describe the longitudinal behaviors of the ADAS-Cog items; to determine how much of the information that they provide is reliable, and to determine what the sources of that reliable information might be.

Here, in mild AD, we sought to characterize the information captured by the ADAS-Cog items in terms of how much of the information is reliable, and how much of that reliable information is consistent with other measurements made over time versus how much is inconsistent (i.e., related only to the specific occasion on which the measurement was made). The consistent information can be further decomposed into information that was already present at the first occasion of measurement (i.e., an individual’s “trait”) versus how much evolved

## RESEARCH IN CONTEXT

- 1. Systematic review:** Previous research showed that items of Alzheimer’s Disease Assessment Scale Cognitive Subscale (ADAS-Cog) behave psychometrically different across time, bringing into question the value of summed scores when used longitudinally.
- 2. Interpretation:** Each ADAS-Cog item captured different amounts of information related to occasion-specific, trait, and accumulated effects of Alzheimer’s disease (AD) over time. The language items were more consistent over time than they were sensitive to occasion-specific changes, and of information that was consistent, the language items tended to reflect effects of AD progression that accumulated from one visit to the next. The study reinforces the assertion that ADAS-Cog items should not be summed to track change over time in mild AD, although some individual items are reliable.
- 3. Future directions:** Because most items were unreliable, and language items were generally the most reliable, over 2 years, longer term studies (e.g., 12–24 months) might consider these items individually to track AD progression. Because only two ADAS-Cog items showed acceptable reliability at all four assessments, word recall (memory) and naming (language), these items might be preferred.

with the progression of AD from one observation to the next (i.e., an “autoregressive” or “accumulated” effect; Figure 1). This study aimed to estimate these features using a latent state-trait autoregressive (LST-AR<sup>21</sup>) model (Figure 1) offering new insight into the interpretation of scores on the ADAS-Cog items in studies of people with mild AD.

## 2 | METHODS

### 2.1 | Data source

Data used in the preparation of this article were obtained from the Alzheimer’s Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). The ADNI was launched in 2003 as a public–private partnership, led by Principal Investigator Michael W. Weiner, MD. The primary goal of ADNI has been to test whether serial magnetic resonance imaging (MRI), positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of MCI and early AD. Participants with mild AD who had ADAS-Cog scores available from at least their baseline visits were included in analyses. Briefly, patients met National Institute of Neurological and Communicative Disorders and Stroke/Alzheimer’s Disease and Related Disorders Association

criteria, a score between 20 and 26 on the Mini-Mental State Examination (MMSE), and a score of 0.5 or 1.0 on the Clinical Dementia Rating Scale. ADNI criteria included AD participants who had changes in memory according to the Logical Memory II subscale of the Revised Wechsler Memory Scale (<http://adni.loni.usc.edu/>). Participants from the ADNI in the cognitively normal or MCI subcohorts were excluded. Data were obtained in January 2018. All participants provided written informed consent and Health Insurance Portability and Accountability Act authorizations were obtained.

## 2.2 | ADAS-Cog protocol, item, and subdomains

The ADAS-Cog was conducted on ADNI participants at their baseline, 6-month, and each annual visit by an Alzheimer's Disease Cooperative Study-ADAS certified psychometrist (see <http://adni.loni.usc.edu/>). Certification is important as some items require subjective evaluation. The 13-item version of the ADAS-Cog (ADAS-Cog 13) was administered to participants. Items assessed in the ADAS-Cog 13 are (1) word recall, (2) commands, (3) constructional praxis, (4) delayed word recall, (5) naming, (6) ideational praxis, (7) orientation, (8) word recognition, (9) remembering test instructions, (10) comprehension of spoken language, (11) word finding difficulty, (12) spoken language ability, and (13) number cancellation. To avoid practice effects,<sup>22</sup> an alternate list of words was used in the "word recall" task (item 1) and recalled in the "delayed word recall" task (item 4) at the 6-month visit. All annual visits (which were considered sufficiently spaced out in time to extinguish practice effects) used the original word list from the baseline visit.

## 2.3 | Statistical analysis

ADAS-Cog data from subjects' baseline, 6-month, 12-month, and 24-month visits were used in analyses. Three LST-AR models with indicator-specific trait variables were analyzed.<sup>21</sup> Latent state-trait (LST) theory is a comprehensive theoretical framework developed to identify and measure sources of behavior variability longitudinally to help address the stability  $\times$  situation debate (for a major details on LST see Steyer et al.<sup>23</sup>). Recently, Eid et al.<sup>21</sup> proposed incorporation of an auto-regressive (AR) feature to acknowledge that traits too can change over time. The LST-AR assumes that behavior is a dynamic system that can change according to situational influences, and that the effects of those influences can accumulate over time. These differences regarding the stability due to trait and due to AR, or accumulated situational effects, are particularly important to allow a better understanding of stability and change processes.

The LST-AR allows for the decomposition of each item at each wave into reliable information (i.e., reliability) and unreliable information (i.e., measurement error); in other words, the LST-AR determines what proportion of the variance of an item score represents reliable information, and what proportion is not reliable. The LST-AR further decomposes the reliable variance of an item into (1) a component that

represents consistent interindividual differences (that are not specific for an occasion of measurement), and (2) a variance component that is due to occasion-specific influences (i.e., dependent on day-to-day fluctuations in performance). Finally, the model decomposes the consistent part of the item variance into (1) a component due to trait effects (i.e., the component of the score that represents an inherent disposition level of cognition that is already present at the baseline assessment), and (2) a component that reflects a source of stability due to "auto-regressive" accumulated occasion-specific effects. To summarize, an observed item score can be decomposed into components:

1. Observed item score = reliable component + measurement error
2. Reliable component = consistent component + occasion-specific effects
3. Consistent component = trait effects + accumulated situational effects

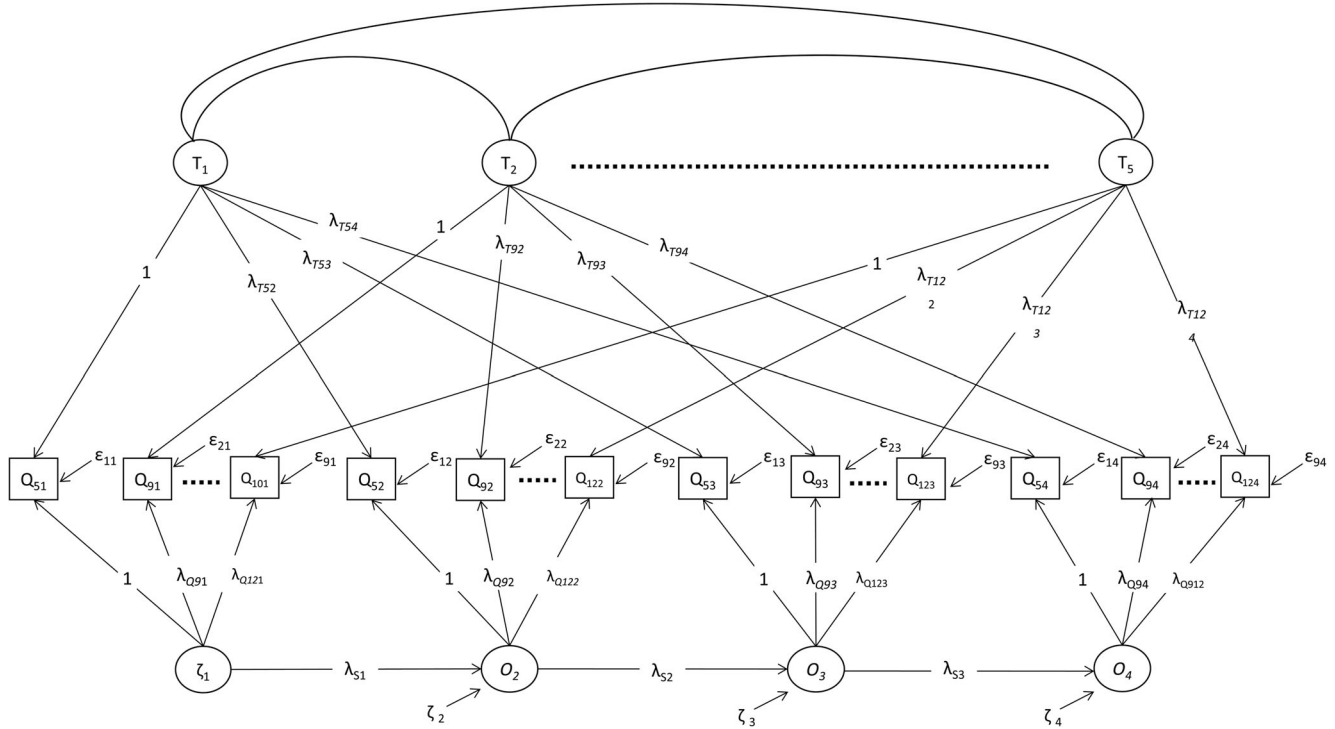
The components are due to different effects:

1. Trait effect: a disposition being predictable from the baseline assessment (i.e., one's inherent level of cognition, in this case)
2. Occasion-specific effects: = due to the situation on an occasion of measurement and/or person-situation interactions
3. AR or accumulated effects: a change or progression that is consistent or predictable over time (e.g., AD-induced cognitive decline)

A visual depiction of the breakdown of the components of a hypothetical assessment score is shown in Figure 1.

For metrical observed items, we specified the model exactly as described previously.<sup>21</sup> This was the case for the four items (word recall, orientation, remembering test instructions, and word recognition) assessing the memory domain (per the ADAS-Cog 11). For ordered categorical observed variables, we applied an LST-AR model.<sup>24</sup> This rationale applies to language subscale constituted by five ordered-categorical items (i.e., commands, naming, comprehension of spoken language, word finding difficulty, and spoken language ability) and praxis constituted by three ordered-categorical items based on the empirical solution of Verma et al.<sup>9</sup> (i.e., commands, constructional praxis, and ideational praxis). Figure 2 depicts the LST-AR model applied to the three subdomains separately.

We impose a measurement invariance restriction on occasion-specific latent variables such that the factor loadings of the items with the same content and occasion-specific latent variables are held equal (e.g., item 5 at baseline will have the same unstandardized factor loading as item 5's factor loading onto the occasion-specific latent variable at times 6 months, 12 months, and 24 months). This measurement invariance assumption implies that the construct does not change over time. This assumption is usually made in longitudinal studies as it simplifies the interpretation of the results and ensures that the same construct is considered over time. The model does not require trait loadings to be the same over time because they represent the influence of the disposition on the first occasion of measurement and this influence can change over time without changing the meaning of



**FIGURE 2** Representation of the latent state-trait autoregressive (LST-AR) statistical model for the language subdomain.  $T_i$  represents trait variables (time-specific dispositions) on the first occasion of measurement, having an influence on the same content item on all occasions of measurement (i.e., naming at baseline, 6 months, 12 months, and 24 months were loaded at  $T_{\text{naming}}$ ). The residuals of the observed ordered-categorical variables are fixed to 1 for a parameterization according to the normal ogive-graded response model. Given sparse data issues (number of thresholds changing across time), thresholds across the different waves were constrained to be equal when available. The occasion-specific latent variable  $O_t$  is composed of  $\zeta_{it}$  (the occasion-specific influences on occasion  $t$ ) and  $\lambda_{s(t-1)} \cdot \zeta_{i(t-1)}$  (indicating carry-over effects or, as called here, the accumulative effects). The correlations between  $T_i$  and the residual variables ( $O_{it}$  and  $\zeta_{it}$ ) were fixed to 0, where  $i$  represents the indicator (i.e., item) and  $t$  the occasion of measurement (i.e., each wave, assessment, sweep). On the first occasions of measurement, the common residual variable is the state residual on the occasion of measurement ( $\zeta_{i1}$ ), whereas subsequently the  $O_t$  variables, the composed-occasion residuals, are the state residuals plus a linear combination of the previous state residuals. Finally, the correlations between the trait variables were freely estimated (i.e., double-headed arrows among all the  $T_i$  latent variables).

the construct. Further constraints can be added if specific hypotheses about the influence of the trait over time should be tested; however, we had no hypotheses about this influence. Because the first two time gaps (i.e., baseline to 6 months, and 6 months to 12 months) were equivalent, we constrained their unstandardized autoregressive effects to be equivalent, leaving the last gap (i.e., 12 months to 24 months) to be estimated freely. We assumed a priori that the autoregressive process would be stable over time (which is typically the case in longitudinal studies). We tested the totality of these assumptions by the confirmatory factor analysis test statistics as an omnibus test to avoid problems of multiple comparisons.

Missing data are expected due to the longitudinal design and depending on the estimator under weighted least square mean and variance adjusted and its pairwise deletion. In the case of the memory domain items, we used the full information maximum likelihood estimator.

Model fit was determined using the  $\chi^2$ -test, comparative fit index (CFI), root mean square error of approximation (RMSEA), and standardized root mean square residual (SRMR), based on the following recommendations for good fit was indicated by a CFI  $\geq 0.97$ , an

RMSEA  $\leq 0.05$  ( $P$ -value  $\leq 0.05$ ), and an SRMR smaller than 0.08.<sup>25</sup> Related to effect size, a cut-off of 70% for reliability was considered to be adequate.<sup>26</sup>

### 3 | RESULTS

#### 3.1 | Sample

In total, 341 subjects with mild AD who had baseline ADAS-Cog scores available were included as reported previously.<sup>20</sup> At baseline, participants were, on average, 75 years old (standard deviation [SD] = 5), 55% male, and had 15 years of education (SD = 3). The percentage of apolipoprotein E  $\epsilon 4$  carriers was 66%. Baseline MMSE score was 23.2 (SD = 2.1), progressing to 18.7 (SD = 5.7) after 24 months.

#### 3.2 | Model fit

All three models (one for each domain) showed appropriate goodness-of-fit indices: memory:  $\chi^2_{(90)} = 117.921$ ,  $P$ -value = 0.0258,

**TABLE 1** Reliabilities of each ADAS-Cog item at each assessment over 24 months.

Subdomain	Item	Reliability			
		Baseline	Month 6	Month 12	Month 24
Memory	Q1	0.859	0.794	0.875	0.853
	Q4	0.719	0.583	0.738	0.808
	Q7	0.470	0.611	0.819	0.531
	Q8	0.419	0.525	0.440	0.435
Language	Q5	0.802	0.817	0.809	0.834
	Q9	0.666	0.784	0.780	0.824
	Q10	0.638	0.745	0.868	0.908
	Q11	0.704	0.698	0.797	0.863
Praxis	Q12	0.686	0.768	0.885	0.925
	Q2	0.434	0.671	0.596	0.801
	Q3	0.544	0.670	0.662	0.749
	Q6	0.555	0.696	0.665	0.767

Abbreviations: ADAS-Cog, Alzheimer's Disease Assessment Scale Cognitive Subscale; Q1–Q12: Q1, word recall; Q2, commands; Q3, construction praxis; Q4, delayed word; Q5, naming; Q6, ideational praxis; Q7, orientation; Q8, word recognition; Q9, remembering the instruction; Q10, comprehension of spoken language; Q11, word finding difficulty; Q12, spoken language ability.

RMSEA = 0.03 (90% confidence interval [CI] = 0.011 to 0.044), *P*-value of RMSEA = 0.992, CFI = 0.988, Tucker–Lewis index (TLI) = 0.983, SRMR = 0.039; language:  $\chi^2_{(201)} = 238.732$ , *P*-value = 0.0353, RMSEA = 0.023 (90% CI = 0.007 to 0.034), *P*-value of RMSEA = 1, CFI = 0.994, TLI = 0.995, SRMR = 0.05; and praxis:  $\chi^2_{(74)} = 92.545$ , *P*-value = 0.0712, RMSEA = 0.027 (90% CI = 0.000 to 0.043), *P*-value of RMSEA = 0.993, CFI = 0.993, TLI = 0.994, SRMR = 0.038. Because all models including the specified restrictions fit the data very well, we did not reject our assumptions about model structure.

### 3.3 | ADAS-Cog item and subdomain reliability

The reliabilities of all items at each time point can be found in Table 1. All language items and two of the memory items (item 1: word recall, item 4: delayed word recall) were generally reliable across most time points. None of the praxis items met the cut-off for adequate reliability across the four time points.

### 3.4 | ADAS-Cog item and subdomain consistency and occasion-specific effects

Of the reliable information captured by each item, a decomposition of trait, accumulated, and state effects are summarized in Table 2. All items had more consistency (trait + accumulated effects) than occasion-specific effects, ranging from 63.4% (item 10) to 88.2% (item 8) indicating that interindividual differences are relatively stable over time.

Visual depiction of the decomposition of variance of the components of the ADAS-Cog assessment at 24 months is shown in Figure 3. Here, measurement error is also depicted.

For three items in the language subdomain (naming, remembering the instructions, word finding difficulty) and three items in the memory subdomain (word recall, delayed word recall, orientation), their consistencies consisted of substantial proportions of trait and accumulated effects. For the remaining language items (comprehension of spoken language and spoken language ability items), accumulated effects comprised most of the consistent information. For the remaining memory item (word recognition) and for all praxis items (commands, construction praxis, and ideational praxis), when splitting consistency into trait versus accumulated effect, most of the consistent information was accounted for by trait effects.

## 4 | DISCUSSION

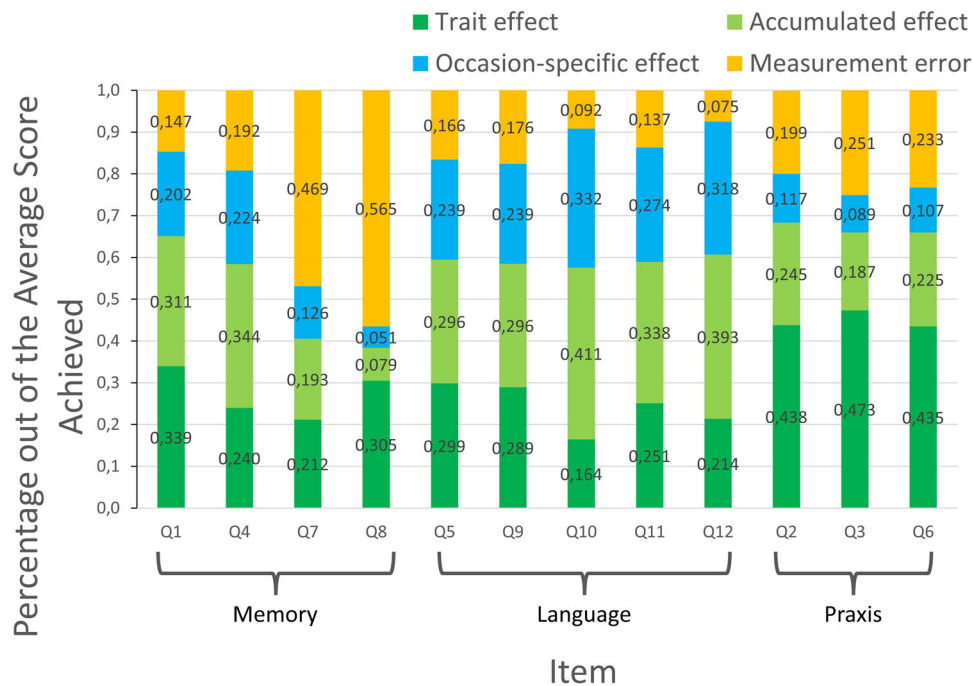
The findings indicated that items from the language subdomain were generally more reliable compared to the other ADAS-Cog subdomains, which tended to contain more measurement error in this mild AD population. Even so, only two items (“word recall” and “naming” from memory and language, respectively) achieved >70% reliability on all four visits over 24 months. These are some of the earlier changes observed in AD. Two items from the memory subdomain (item 1 “word recall” and item 4 “delayed word recall”) and all language items exhibited acceptable reliabilities on at least three occasions. Considering decomposition into occasion-specific, trait, and accumulated effects, these items from the language and memory subdomains may capture AD-related cognitive changes over time most effectively, as they were most highly influenced by accumulated effects (vs. occasion-specific or trait effects). Previous psychometric network analysis showed that item 10 “comprehension of spoken language” appeared to be the most sensitive to psychopharmacological intervention, suggesting that it might be

**TABLE 2** Consistency (trait and accumulated effects) and occasion-specific effects decomposed for each ADAS-Cog item at the 24-month assessment.

Subdomain	Item	Consistency			Inconsistency Occasion-specific effects
		Total consistency	Trait effects	Accumulated effects	
Memory	Q1	0.763	0.398	0.365	0.237
	Q4	0.723	0.297	0.426	0.277
	Q7	0.763	0.399	0.364	0.237
	Q8	0.882	0.701	0.181	0.118
Language	Q5	0.713	0.358	0.355	0.287
	Q9	0.710	0.351	0.359	0.290
	Q10	0.634	0.181	0.453	0.366
	Q11	0.683	0.291	0.392	0.317
	Q12	0.656	0.231	0.425	0.344
Praxis	Q2	0.854	0.547	0.306	0.146
	Q3	0.881	0.632	0.249	0.119
	Q6	0.860	0.567	0.293	0.140

Note: Taking reliability as constituted by 100% of trustworthy variance, the reliable information for each item at every time, as per Figure 2, might be disentangled in two parts (consistency [trait effect + accumulated effect] and inconsistency).

Abbreviations: ADAS-Cog, Alzheimer's Disease Assessment Scale Cognitive Subscale; Q1, word recall; Q2, commands; Q3, construction praxis; Q4, delayed word; Q5, naming; Q6, ideational praxis; Q7, orientation; Q8, word recognition; Q9, remembering the instruction; Q10, comprehension of spoken language; Q11, word finding difficulty; Q12, spoken language ability.



**FIGURE 3** Visual depiction of the decomposition of variance of the components of the Alzheimer's Disease Assessment Scale Cognitive Subscale assessment at 24 months. All the sources of variance are described here: consistency [trait effect + accumulated effect], inconsistency [occasion-specific effect], and measurement error. All sources together sum to 100%. Q1, word recall; Q2, commands; Q3, constructional praxis; Q4, delayed word recall; Q5, naming; Q6, ideational praxis; Q7, orientation; Q8, word recognition; Q9, remembering test instructions; Q10, comprehension of spoken language; Q11, word finding difficulty; Q12, spoken language ability.

the most important to monitor among outcomes.<sup>27</sup> This item is a subjective observation made by the examiner, so high reliability might not have been anticipated. Nonetheless, our results concur, and our approach provides two possible theoretical bases for that finding. First, “comprehension of spoken language”, along with “word recall” and “naming”, was found to be generally reliable, having a relatively low proportion of measurement error, which would make it among the items more sensitive to participant performance. Second, when examining the behavior of the items over time, “comprehension of spoken language” was found to have the highest proportion of occasion-specific effect, the highest proportion of accumulated effects, and the lowest proportion of trait effects, of any ADAS-Cog item. Sensitivity to occasion-specific and/or accumulated effects may indicate that the “comprehension of spoken language” item tracked the accumulation of AD-related deficits over time (i.e. accumulated effect), and/or the true variability in performance from visit to visit (i.e., state), to the greatest extent of the ADAS-Cog items. However, the context for these observations may be important, as for instance, the baseline severity and inclusion/exclusion criteria for the cohort studied (discussed below).

All items in the praxis subdomain (“commands”, “constructional praxis”, and “ideational praxis”), and two items in the memory subdomain (“orientation” and “word recognition”) had relatively lower reliabilities, which would not meet a general threshold of  $\geq 70\%$ . If a memory estimate was required from existing study data, the present findings would suggest the preferred use of items 1 and 4 (“word recall” and “delayed word recall”). The praxis subdomain items were subject to considerable measurement error, and therefore the ways in which the ADAS-Cog assesses praxis may be unreliable, suggesting that if needed, an alternative assessment of praxis might be implemented going forward.

Clinical trials for AD therapeutics have commonly used some version of the ADAS-Cog at longitudinal visits to assess the efficacy of the treatment in slowing cognitive decline; for nearly two decades, trials of amyloid beta ( $A\beta$ )-targeted therapies have failed to show efficacy,<sup>28</sup> although some have shown results in the ADAS-Cog that were trending toward significance.<sup>29</sup> If language item scores are reliable and most heavily influenced by the progression of mild AD, the clinical trial data might be reassessed post hoc using language subdomain items, or specifically “comprehension of spoken language,” as most items from the other subdomains were less reliable, and their reliable information was more heavily related to interindividual trait differences. This may have diluted the efficacy measurement, as suggested empirically in the context of trials by Rotstein et al.<sup>27</sup> Memory scores, in particular the items “word recall” and “delayed word,” were found also to be fairly reliable with similar contributions of accumulated effects to the reliable, consistent information that they yield. In contrast, studies of the cholinesterase inhibitors showed effects on the ADAS-Cog; however, treatment efficacy is known to wane over extended timelines such as those in this study, suggesting that those shorter term changes on the ADAS-Cog may not be relevant to longer term decline described here. It remains to be explored how alternative cognitive measures might perform under state–trait models as, for example, LST-AR and their

derived indices. Psychometric studies have compared the ADAS-Cog to alternatives such as the Neuropsychological Test Battery, finding superior characteristics, but no study has explored state–trait features of those alternatives.<sup>14</sup> It is unclear if the characteristics uncovered in this study are specific to the ADAS-Cog items, and similar models might be applied to other tests, including newer computerized batteries.

It is possible that the inclusion/exclusion criteria used for the mild AD group in this study biased the memory scores to show greater trait effects relative to accumulated effects; inclusion criteria required changes in memory, which meant that included participants were biased to have lower levels of memory but not necessarily other changes, which may have affected state, trait, and accumulated estimates over time. Nonetheless, the data are still likely relevant to the interpretation of trials, since they have often used inclusion criteria similar or identical to those of the ADNI. Different aspects of AD symptoms may progress differently at different stages of the disease, and therefore they may exhibit different occasion-specific and trait features at different stages of the disease, which should be examined in future studies. Here, MMSE scores progressed from 23.2 to 18.7 over the 24-month observation period, indicating that cognitive decline had occurred; however, the specific state–trait results may not generalize to other stages of AD progression. The LST-AR-derived indices should be compared between this and other samples at similar stages of decline, and explored in other samples at different stages of decline, to determine generalizability.

More recently, the field has begun to shift toward a biological definition, including biomarker measures under the ATN framework (i.e., levels of  $A\beta$ , tau, and neurodegeneration using fluid biomarkers, PET scans, or MRI).<sup>30,31</sup> It is unclear how the incorporation of these biomarker measures as inclusion criteria will affect the longitudinal characteristics of the outcome measures, particularly if trial participants continue to be identified by cognitive impairment prior to applying biomarker criteria; however, it may be important to consider in these new contexts, in a fashion similar to that of the present work, the proportions of variance in outcome measures that were reliable, and of that information how much was immovable versus accumulated or fluctuating unpredictably between visits. This may help to characterize the outcome measures and to identify and mitigate biases that could dampen their ability to track cognitive decline effectively. In particular, the trait information may be immovable, and in the context of trials, that component should be evened out between the groups by the randomization process; the LST-AR models, particularly under multi-group structures, could be used to test this explicitly.

## 5 | CONCLUSION

The current study provides insight into the interpretation and utility of the ADAS-Cog in mild AD. A significant proportion of the variance in ADAS-Cog item scores was unreliable in this sample. The five language items, and memory items “word recall” and “delayed word”, were the most reliable, and they most effectively (35%–45%) measured the



accumulation of cognitive deficits and/or the reliable visit-to-visit fluctuations in cognitive performance (24%–37%). If much of the variance in the ADAS-Cog was unreliable, or reflective of between-subjects trait differences, it may have been unnecessarily difficult to obtain treatment effects using many ADAS-Cog items in clinical trials, and data from previous trials might be reconsidered. Similarly, correlative studies assessing relationships between pathophysiological elements of AD and ADAS-Cog scores might benefit from selecting items with longitudinal characteristics that best match the hypothesis vis-à-vis their occasion-specific, trait, or accumulating effects of AD progression over time. More broadly, the application of state and trait models to other outcome measures may help to identify appropriately reliable and sensitive outcomes.

## ACKNOWLEDGMENTS

Hugo Cogo-Moreira is thankful to CAPES/Alexander von Humboldt Fellowship. Walter Swardfager gratefully acknowledges funding from the Canadian Institutes of Health Research, the Alzheimer's Association (US), Brain Canada, The Michael J. Fox Foundation, Weston Brain Institute, and Alzheimer's Research UK. This research was undertaken, in part, thanks to funding from the Canada Research Chairs Program (Walter Swardfager). Jennifer S. Rabin gratefully acknowledges funding from the Canadian Institutes of Health Research (173253, 438475), and the Alzheimer's Society of Canada. Krista L. Lanctôt acknowledges the Bernick Chair in Geriatric Psychopharmacology.

Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie; Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC; Johnson & Johnson Pharmaceutical Research & Development LLC; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health ([www.fnih.org](http://www.fnih.org)). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Therapeutic Research Institute at the University of Southern California. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California. Funding was provided by the Alzheimer's Association (USA) and Brain Canada.

## CONFLICTS OF INTEREST

All authors have no actual or potential competing of interests to declares. Author disclosures are available in the [supporting information](#).

## REFERENCES

1. Kueper JK, Speechley M, Montero-Odasso M. The Alzheimer's Disease Assessment Scale-Cognitive Subscale (ADAS-Cog): modifications and responsiveness in pre-dementia populations. A narrative review. *J Alzheimers Dis*. 2018;63(2):423-444.
2. Rosen WG, Mohs RC, Davis KL. A new rating scale for Alzheimer's disease. *Am J Psychiatry*. 1984;141(11):1356-1364.
3. Bengtson JF, Balsis S, Geraci L, Massman PJ, Doody RS. How well do the ADAS-cog and its subscales measure cognitive dysfunction in Alzheimer's disease? *Dement Geriatr Cogn Disord* 2009;28(1):63-69.
4. Schrag A, Schott JM; Alzheimer's Disease Neuroimaging I. What is the clinically relevant change on the ADAS-Cog? *J Neurol Neurosurg Psychiatry*. 2012;83(2):171-173.
5. Mohs RC, Knopman D, Petersen RC, et al. Development of cognitive instruments for use in clinical trials of antidementia drugs: additions to the Alzheimer's Disease Assessment Scale that broaden its scope. The Alzheimer's Disease Cooperative Study. *Alzheimer Dis Assoc Disord*. 1997;11(suppl 2):S13-S21.
6. Skinner J, Carvalho JO, Potter GG, et al. The Alzheimer's Disease Assessment Scale-Cognitive-Plus (ADAS-Cog-Plus): an expansion of the ADAS-Cog to improve responsiveness in MCI. *Brain Imaging Behav*. 2012;6(4):489-501.
7. Llano DA, Laforet G, Devanarayan V; Alzheimer's Disease Neuroimaging I. Derivation of a new ADAS-cog composite using tree-based multivariate analysis: prediction of conversion from mild cognitive impairment to Alzheimer disease. *Alzheimer Dis Assoc Disord* 2011;25(1):73-84.
8. Raghavan N, Samtani MN, Farnum M, et al. The ADAS-Cog revisited: novel composite scales based on ADAS-Cog to improve efficiency in MCI and early AD trials. *Alzheimers Dement*. 2013;9(1 Suppl):S21-S31.
9. Verma N, Beretvas SN, Pascual B, Masdeu JC, Markey MK, Alzheimer's Disease Neuroimaging I. New scoring methodology improves the sensitivity of the Alzheimer's Disease Assessment Scale-Cognitive subscale (ADAS-Cog) in clinical trials. *Alzheimers Res Ther*. 2015;7(1):64.
10. Grochowalski JH, Liu Y, Siedlecki KL. Examining the reliability of ADAS-Cog change scores. *Neuropsychol Dev Cogn B Aging Neuropsychol Cogn*. 2016;23(5):513-529.
11. Rockwood K, Fay S, Gorman M, Carver D, Graham JE. The clinical meaningfulness of ADAS-Cog changes in Alzheimer's disease patients treated with donepezil in an open-label trial. *BMC Neurol*. 2007;7:26.
12. Cano SJ, Posner HB, Moline ML, et al. The ADAS-cog in Alzheimer's disease clinical trials: psychometric evaluation of the sum and its parts. *J Neurol Neurosurg Psychiatry*. 2010;81(12):1363-1368.
13. Ueckert S, Plan EL, Ito K, et al. Improved utilization of ADAS-cog assessment data through item response theory based pharmacometric modeling. *Pharm Res*. 2014;31(8):2152-2165.
14. Karin A, Hannesdottir K, Jaeger J, et al. Psychometric evaluation of ADAS-Cog and NTB for measuring drug response. *Acta Neurol Scand*. 2014;129(2):114-122.
15. Kottner J, Dassen T. An interrater reliability study of the Braden scale in two nursing homes. *Int J Nurs Stud* 2008;45(10):1501-1511.
16. Terwee CB, Bot SD, de Boer MR, et al. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol*. 2007;60(1):34-42.
17. Shoukri MM, Asyali M, Donner A. Sample size requirements for the design of reliability study: review and new results. *Statis Methods Med Res*. 2004;13(4):251-271.

18. Polit DF, Beck CT. *Nursing Research: Generating and Assessing Evidence for Nursing Practice*. 8th ed. Lippincott Williams & Wilkins; 2008.
19. Nunnally JC, Bernstein IH. *Psychometric Theory*. 3rd ed. McGraw-Hill; 1994.
20. Cogo-Moreira H, Krance SH, Black SE, et al. Questioning the meaning of a change on the Alzheimer's Disease Assessment Scale-Cognitive Subscale (ADAS-Cog): noncomparable scores and item-specific effects over time. *Assessment*. 2021;28(6):1708-22.
21. Eid M, Holtmann J, Santangelo P, Ebner-Priemer U. On the definition of latent-state-trait models with autoregressive effects. *Eur J Psychol Assess*. 2017;33(4):285-295.
22. Jacobs DM, Ard MC, Salmon DP, Galasko DR, Bondi MW, Edland SD. Potential implications of practice effects in Alzheimer's disease prevention trials. *Alzheimers Dement (N Y)*. 2017;3(4):531-535.
23. Steyer R, Mayer A, Geiser C, Cole DA. A theory of states and traits-revised. *Annu Rev Clin Psychol*. 2015;11:71-98.
24. Eid M. Longitudinal confirmatory factor analysis for polytomous item responses: model definition and model selection on the basis of stochastic measurement Theory. *Methods Psychol Res*. 1996;1:65-85.
25. Schermelleh-Engel K, Moosbrugger H, Müller H. Evaluating the fit of structural equation models: tests of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research*. 2003;8(2):23-74.
26. Cortina JM. What is coefficient alpha? an examination of theory and applications. *Journal of Applied Psychology*. 1993;78(1):98-104.
27. Rotstein A, Levine SZ, Samara M, et al. Cognitive impairment networks in Alzheimer's disease: analysis of three double-blind randomized, placebo-controlled, clinical trials of donepezil. *Eur Neuropsychol*. 2022;57:50-58.
28. Panza F, Lozupone M, Logroscino G, Imbimbo BP. A critical appraisal of amyloid-beta-targeting therapies for Alzheimer disease. *Nat Rev Neurol*. 2019;15(2):73-88.
29. Honig LS, Vellas B, Woodward M, et al. Trial of solanezumab for mild dementia due to Alzheimer's disease. *N Engl J Med*. 2018;378(4):321-330.
30. Cummings J. The national institute on aging-Alzheimer's association framework on Alzheimer's disease: application to clinical trials. *Alzheimers Dement*. 2019;15(1):172-178.
31. Jack CR Jr., Bennett DA, Blennow K, et al. NIA-AA research framework: toward a biological definition of Alzheimer's disease. *Alzheimers Dement*. 2018;14(4):535-562.

## SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

**How to cite this article:** Cogo-Moreira H, Krance SH, Wu C-Y, et al. State, trait, and accumulated features of the Alzheimer's Disease Assessment Scale Cognitive Subscale (ADAS-Cog) in mild Alzheimer's disease. *Alzheimer's Dement*. 2023;9:e12376. <https://doi.org/10.1002/trc2.12376>