Received: 7 November 2023 | Revised: 28 March 2024 | Accepted: 30 March 2024

DOI: 10.1002/trc2.12471

Alzheimer's & Dementia
Translational Research & Clinical Interventions

RESEARCH ARTICLE

Assessing tilavonemab efficacy in early Alzheimer's disease via longitudinal item response theory modeling

Xiaoxiao Zhou[1] | **Haotian Zou**[1] | **Michael W. Lutz**[2] | **Konstantin Arbeev**[3] | **Igor Akushevich**[3] | **Anatoli Yashin**[3] | **Kathleen A. Welsh-Bohmer**[4,5] | **Sheng Luo**[1]

[1]Department of Biostatistics & Bioinformatics, Duke University, Durham, North Carolina, USA

[2]Division of Translational Brain Sciences, Department of Neurology, Duke University Medical Center, Durham, North Carolina, USA

[3]Social Science Research Institute, Duke University, Durham, North Carolina, USA

[4]Department of Psychiatry, Duke University, Durham, North Carolina, USA

[5]Duke Clinical Research Institute (DCRI), Duke University, Durham, North Carolina, USA

Correspondence
Sheng Luo, Department of Biostatistics & Bioinformatics, Duke University, Durham, NC, USA.
Email: sheng.luo@duke.edu

Funding information
National Institute on Aging, Grant/Award Numbers: AG064803, P30AG072958

Abstract

INTRODUCTION: Alzheimer's disease (AD) is a neurodegenerative disorder characterized by declines in cognitive and functional severities. This research utilized the Clinical Dementia Rating (CDR) to assess the influence of tilavonemab on these deteriorations.

METHODS: Longitudinal Item Response Theory (IRT) models were employed to analyze CDR domains in early-stage AD patients. Both unidimensional and multidimensional models were contrasted to elucidate the trajectories of cognitive and functional severities.

RESULTS: We observed significant temporal increases in both cognitive and functional severities, with the cognitive severity deteriorating at a quicker rate. Tilavonemab did not demonstrate a statistically significant effect on the progression in either severity. Furthermore, a significant positive association was identified between the baselines and progression rates of both severities.

DISCUSSION: While tilavonemab failed to mitigate impairment progression, our multidimensional IRT analysis illuminated the interconnected progression of cognitive and functional declines in AD, suggesting a comprehensive perspective on disease trajectories.

KEYWORDS
bayesian modeling, clinimetrics, disease progression, longitudinal data

Highlights

1. Utilized longitudinal Item Response Theory (IRT) models to analyze the Clinical Dementia Rating (CDR) domains in early-stage Alzheimer's disease (AD) patients, comparing unidimensional and multidimensional models.
2. Observed significant temporal increases in both cognitive and functional severities, with cognitive severity deteriorating at a faster rate, while tilavonemab showed no statistically significant effect on either domain's progression.

3. Found a significant positive association between the baseline severities and their progression rates, indicating interconnected progression patterns of cognitive and functional declines in AD.
4. Introduced the application of multidimensional longitudinal IRT models to provide a comprehensive perspective on the trajectories of cognitive and functional severities in early AD, suggesting new avenues for future research including the inclusion of time-dependent random effects and data-driven IRT models.

1 | INTRODUCTION

Alzheimer's disease (AD) is a leading neurodegenerative disorder leading to cognitive and functional impairments and is the principal cause of dementia worldwide.[1] Its complex consequences span emotional, financial, and societal dimensions. Given the main therapeutic emphasis on addressing the inherent pathology of AD, the Clinical Dementia Rating Scale (CDR) has become an essential instrument for assessing dementia severity. The CDR categorizes the clinical expression of dementia into six domains: three cognitive—Memory, Orientation, and Judgment and Problem Solving—and three functional—Community Affairs, Home and Hobbies, and Personal Care.[2] Clinicians assign scores based on patient interviews and collateral sources like caregivers. For the first five domains, scores range from 0 (no impairment) to 3 (severe impairment), with intermediate scores of 0.5 (questionable), 1 (mild), and 2 (moderate). The Personal Care domain has a similar scale but typically excludes the 0.5 score.[3]

The CDR commonly utilizes the sum of the boxes (CDR-SB) scoring method, summing scores across its six domains with the total CDR-SB score ranging from 0 to 18. Recognizing the dual cognitive and functional concepts captured by the CDR, the FDA (U.S. Food and Drug Administration) recommended its use in early symptomatic, stage 3 AD[4] and as a co-primary endpoint in mild-to-moderate AD dementia.[5] CDR-SB has been adopted in many AD disease-modifying therapy trials, for example, studies of aducanumab,[6] lecanemab,[7] and tilavonemab.[8] Typically, longitudinal CDR-SB scores from these studies are treated as continuous variables, analyzed using mixed models or generalized estimating equations (GEE). However, this methodology assumes a unidimensional representation of the disease, equal significance and sensitivity of each domain, and identical domain discrimination. These assumptions might not be valid, particularly when disease progression or treatment effects vary between domains of impairment. Highlighting such concerns, Jutten et al.[9] emphasized the potential risks of composite scoring in missing subtle cognitive variations, while another study found the CDR-SB suboptimal for assessing the efficacy of treatments in slowing disease progression.[10]

Exploring alternate methodologies, Item Response Theory (IRT) modeling offers an advanced approach for analyzing longitudinal CDR data, linking individual item responses to the latent disease severity termed "theta." While IRT has distinct advantages over sum score methods such as CDR-SB, its use in AD research is limited. Typically, studies adopt a unidimensional approach by treating all items as measuring a single attribute of the disease or emphasize cross-sectional data. For example, Vandemeulebroecke *et al*[11] and Bascoul-Mollevi et al.[12] demonstrated the strength of unidimensional longitudinal IRT models using various neuropsychological test items. On the other hand, using cross-sectional IRT modeling, Lowe et al.[3] categorized the six CDR domain scores into a single latent variable while Li et al.[13] systematically evaluate each domain at baseline to identify domains that are most diagnostic of dementia severity.

While IRT modeling is recognized in neurodegenerative research, the application of multidimensional longitudinal IRT models incorporating multiple latent variables[14–19] remains limited in discerning distinct AD progression patterns and associated treatment responses in multiple correlated severities. This study employs these IRT models to analyze AD progression through domain-level CDR responses. As the CDR specifically classifies dementia severity into cognitive and functional severities, our multivariate longitudinal IRT approach offers a thorough perspective on AD's progression over time. With the support of longitudinal data, this methodology models the trajectory of cognitive and functional decline in AD. By reevaluating CDR scores from the recently published AWARE study, we highlight the insights IRT modeling provides, particularly when traditional mixed modeling of CDR-SB indicated no treatment effect.[8]

2 | METHODS

2.1 | Study population

We sourced our data from the phase 2, multiple-dose, double-blind, placebo-controlled AWARE trial of Stage 3 early AD (Clinicaltrials.gov identifier: NCT02880956) conducted across 60 sites in 11 countries. Detailed clinical eligibility criteria for participants can be found in Florian et al.[8] Participants were randomized in equal proportions to one of three tilavonemab doses (300 mg, 1000 mg, 2000 mg) or a placebo, with intravenous infusions administered every 4 weeks from the initial visit. This double-blind treatment spanned 96 weeks, with CDR score evaluations at baseline and subsequently every 24 weeks. For rigorous data analysis, we maintained stringent inclusion and exclusion criteria,

focusing only on participants with complete CDR global scores across their visits. Our data handling yielded an analysis dataset comprising 314 subjects and 1,570 visits, narrowed down from an original cohort of 453 subjects and 2,265 visits. We utilized Locally Weighted Scatterplot Smoothing (LOWESS) curves to smoothly display the trends in cognitive subsum (the sum of three cognitive domains) and functional subsum (the sum of three functional domains). This method draws smooth trends through data by making the line's smoothness adjustable, prioritizes closer points, and iteratively refines to spotlight key patterns.[20]

2.2 | Unidimensional longitudinal IRT model

We utilized a unidimensional longitudinal IRT model to individually assess CDR domains in the cognitive and functional severities. The unidimensional analysis assumed that there is a single latent variable, "theta", in IRT models, representing either the underlying cognitive severity when applied to three cognitive domains (Memory, Orientation, and Judgement and Problem Solving), or the underlying functional severity when applied to three functional domains (Community Affairs, Home and Hobbies, and Personal Care). The unidimensional longitudinal IRT model consists of two levels. The first level, a graded-response measurement model (refer to Figures S1 and S2; Model 1 in Supplemental Materials), quantifies the relationship between each domain's response and the latent variable "theta". The probability of every score in each domain was determined by five parameters: the discrimination parameter and the location parameters. The discrimination parameter gauges the domain's ability to differentiate based on latent severity, with a higher value indicating that this domain is powerful for determining the individual's latent severity. The location parameters (also called difficulty parameters) are the probability threshold for transitioning from score 0 to 0.5, from 0.5 to 1, etc. The second-level structural model (Model 2 in the Supplemental Materials) regresses cognitive or functional severity, indicated by "theta", on time, treatment group, and its interaction with time, with subject-specific random intercepts, and random slopes of time. The random intercepts allow each individual to have a personalized severity at baseline, whereas the random slopes account for individuals' unique progression rates in severity. The correlation between the random intercepts and random slopes was modeled by a correlation coefficient. Because the IRT models are overparameterized, certain constraints need to be set to make the models identifiable (estimable).[21] Specifically, random intercepts were assumed to have standard normal distribution with mean zero and variance one, and the random slopes were assumed to follow a normal distribution with mean zero. The detailed illustrations of the unidimensional longitudinal IRT model can be found in our prior work.[14–19]

We investigated the progression of cognitive and functional severities and compared the effects of tilavonemab with placebo. Assessment of significant impacts relied on 95% credible intervals (Bayesian equivalence to 95% credible interval [CI]). To facilitate the parameter estimation, groups with fewer than 10 subjects were merged with adjacent groups to ensure adequate sample size for analysis. For the Memory domain, scores of 0 and 0.5 were merged into 0.5, and for all other domains, scores of 2 and 3 were unified as 2, leading to adjusted score categories: memory (0.5, 1, 2), personal care (0, 1, 2), and other domains (0, 0.5, 1, 2).

RESEARCH IN CONTEXT

1. **Systematic Review**: A comprehensive literature review, including primary sources such as PubMed and conference presentations, highlighted that while Item Response Theory (IRT) modeling is prevalent in Alzheimer's disease (AD) research, its application has been primarily for cross-sectional analyses and is based on the unidimensional assumption that a single latent variable represents disease severity.
2. **Interpretation**: In early AD, cognitive decline outpaces functional impairment. The trajectories of both severities are intercorrelated, with tilavonemab showing no significant modulation of these progressions.
3. **Future Direction**: This study introduces the application of multidimensional longitudinal IRT models to dissect the interrelated yet distinct trajectories of cognitive and functional severities in early AD. Future research avenues may include: (a) considering the inclusion of time-dependent random effects, and (b) adopting data-driven IRT models to enhance the specificity of understanding the cognitive and functional impairments in AD progression.

2.3 | Multidimensional longitudinal IRT model

The multidimensional longitudinal IRT model allows for multiple latent variables. Given prior CDR knowledge, we posited two separate but correlated latent variables: $theta^{cog}$ for cognitive severity (from three cognitive domains) and $theta^{fun}$ for functional severity (from three functional domains). The first-level graded-response measurement models (refer to Models 5 and 6 in Supplemental Materials) are similar to the unidimensional longitudinal IRT model and they relate each item's response to its pertinent latent variable. In the second-level structural models, both latent variables $theta^{cog}$ and $theta^{fun}$ were regressed on time, treatment group, and its interaction with time, in addition to subject-specific random intercepts and random slopes of time to allow each individual to have his/her baseline severity and rates of progression in both cognitive and functional severities. To account for the relationship between the two *thetas*, the correlation among the random intercepts and random slopes in both impairments was represented in a correlation matrix. Because IRT models are overparameterized, additional constraints need to be imposed to make the

models identifiable (estimable). Specifically, random intercepts U_{i0} and U_{i2}, as in Models 5 and 6 in Supplemental Materials, were assumed to follow a standard normal distribution marginally with mean zero and variance one. This assumption ensures that, for subjects in the placebo group, the multivariate latent variables $theta^{cog}$ and $theta^{fun}$ had a mean zero at baseline and variance one for both severities. The detailed illustrations of the multidimensional longitudinal IRT model can be found in our prior work.[14–19]

We employed the multidimensional longitudinal IRT model to assess the progression of cognitive and functional severities and to contrast the effects of tilavonemab with placebo. A distinct advantage of this model is its unitless latent variables, or theta components, which enable straightforward comparisons and evaluations of progression rates and tilavonemab responsiveness across both cognitive and functional severities. This granularity offers insights that the unidimensional IRT model does not capture. Both models presume a linear relationship between the latent variable theta and time, with the progression patterns illustrated in Figure 1 (lower panels). As sensitivity analysis, we also implemented a linear mixed model to analyze the CDR sum scores, regressing the CDR-SB scores on time, the interaction between treatment group and time, and subject-specific random intercepts and random slopes.

2.4 | Bayesian inference model fitting

We conducted our analyses using Bayesian inference via Markov Chain Monte Carlo (MCMC) posterior simulations in *Stan* (version 2.28)[22] through *R* (version 4.3.1). All model parameters were assigned vague priors, with specifics on prior distribution choices, parameter settings, and convergence assessments detailed in our prior work.[15] We utilized the Deviance information criterion (DIC)[23] for model selection, with lower values indicating a better data fit. The models with smaller DIC are preferred. To enhance the clinical interpretation of the regression coefficients, we applied simulation techniques to express them in terms of CDR domain scores. These simulation techniques utilized the posterior samples from Bayesian inference and were associated with increased variance. Consequently, we expressed the primary statistical results in unitless theta values and provided the simulation CDR domain-based results in the Supplemental Materials. For *Stan* codes, refer to Supplementary Section 3.

2.5 | Model performance diagnostics

The longitudinal IRT models' performance was comprehensively assessed using diagnostics based on simulation and residual evaluations. Using estimated parameters, we simulated responses across all categories of CDR scores for each visit. These simulated values were then directly compared to the observed proportions, with the results visualized through mirror and scatter plots for clearer comparative insights. Residuals for each domain were computed as deviations between observed domain values and weighted prediction scores, derived by aggregating simulated probabilities for category scores. These residuals informed the model's observational value accuracy. We further derived the cognitive and functional subsums and total sum from simulated data, comparing these with observed data to assess model ability in replicating observed sum scores.

3 | RESULTS

3.1 | Baseline characteristics

Table 1 summarizes the baseline characteristics of the analysis dataset, highlighting balanced participant numbers across treatment groups. At baseline, no notable differences were observed in various cognitive metrics, including CDR-SB. The analysis of cumulative cognitive and functional subsum scores, visualized through LOWESS curves in Figure 1 (upper panels), indicates a steady progression of cognitive and functional deterioration over the study period.

3.2 | Model comparison and selection

The multidimensional IRT model demonstrated a better fit (DIC = 12,293) compared to the combined unidimensional models (DIC = 12,516). This highlights the enhanced accuracy of the multidimensional model over the unidimensional approach that separates domains into two distinct latent variables. Moreover, the unidimensional model, encompassing a singular latent variable spanning all six domains, showed substantial coefficient variance (details not included). This suggests the potential inadequacy of assuming a single latent variable for all domains, which aligns with the dual latent variable structure of CDR. In light of these findings and the inherent bidimensional nature of the CDR, we deemed the multidimensional IRT model with the interaction term as our model of choice. Subsequent results discussed are derived from this chosen model, though we also present outcomes from the separate unidimensional models for comparation.

3.3 | Separate unidimensional longitudinal IRT models for cognitive and functional severities

Using the separate unidimensional longitudinal IRT models with three cognitive and three functional domain scores, we identified a significant time effect characterized by a cognitive decline rate of 0.794 theta units/year (95% CI: [0.588, 1.004]) and a functional decline rate of 0.592 theta units/year (95% CI: [0.403, 0.796]). While tilavonemab did not significantly impact cognitive severity across dosages, the 2000 mg dose notably accelerated the functional decline by 0.283 theta units/year (95% CI: [0.033, 0.542]). The other doses showed no significant effect. Furthermore, no significant correlation was found between baseline abilities and decline rates in both severities, indicating baseline severity did not dictate the progression rate. Regression parameters are detailed in Table 2, with location and discrimination

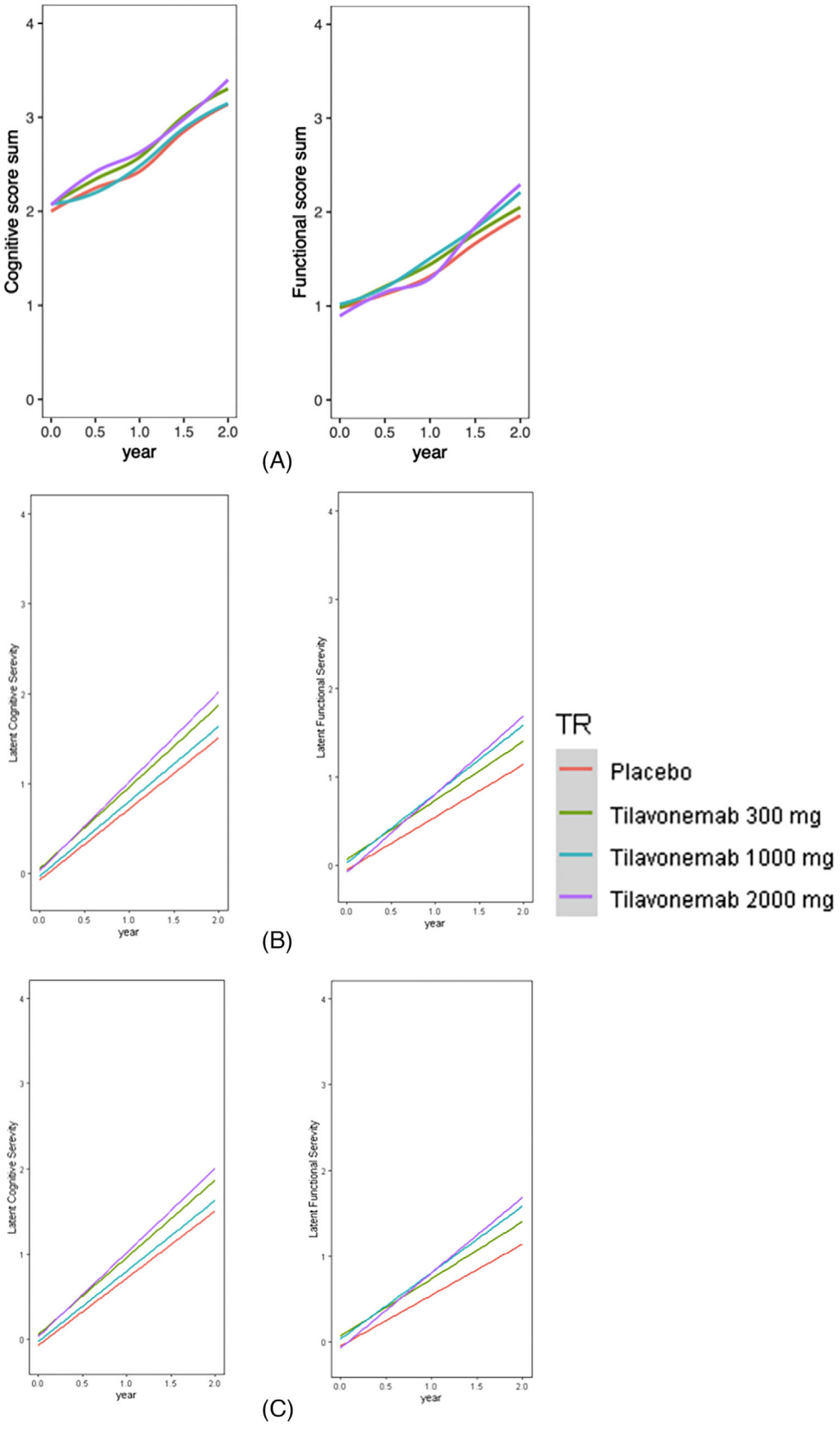

FIGURE 1 Upper panels: LOWESS curves of cognitive subsum and functional subsum in CDR score based on 314 subjects and 1,570 visits. Middle panels: LOWESS curves for latent cognitive and functional severities from two separate unidimensional longitudinal IRT models; Lower panels: LOWESS curves for latent cognitive and functional severities from the multidimensional longitudinal IRT model

TABLE 1 Baseline characteristics of the analysis dataset with 314 subjects and 1,570 visits

Characteristic	Placebo ($n = 79$)	TN 300 mg ($n = 79$)	TN1000 mg ($n = 81$)	TN 2000 mg ($n = 75$)	Total ($n = 314$)
Age (y)	71.7 ± 6.8	70.9 ± 6.7	71.7 ± 7.1	70.0 ± 7.1	71.1 ± 6.9
Female	48 (60.8%)	36 (45.6%)	43 (53.1%)	43 (57.3%)	170 (54.1%)
Race					
White	76 (96.2%)	74 (93.7%)	78 (96.3%)	75 (100.0%)	303 (96.5%)
Non-White	3 (3.8%)	5 (6.3%)	3 (3.7%)	0 (0.0%)	11 (3.5%)
Education (y)	14.7 ± 3.5	15.2 ± 4.1	14.3 ± 4.0	14.8 ± 3.6	14.8 ± 3.3
CDR-SB	3.0 ± 1.3	3.1 ± 1.1	3.1 ± 1.2	3.0 ± 1.0	3.0 ± 1.2
ADAS-Cog14	26.1 ± 7.8[a]	27.1 ± 7.9[b]	25.7 ± 7.5	26.5 ± 7.2[c]	26.3 ± 7.6[d]
RBANS	72.5 ± 11.0[e]	72.5 ± 14.4[f]	72.3 ± 12.5	70.1 ± 12.6	71.9 ± 12.7[g]
MMSE	24.2 ± 3.0	24.5 ± 2.9	24.5 ± 2.7	24.6 ± 3.0	24.4 ± 2.9
FAQ	7.3 ± 5.6	7.7 ± 5.4	8.4 ± 5.4	7.7 ± 4.7	7.8 ± 5.3
NPI	3.4 ± 4.6	3.7 ± 5.2	4.2 ± 6.2	3.5 ± 5.2	3.7 ± 5.3
UPSA-Brief	60.3 ± 19.3	61.5 ± 18.9	59.7 ± 18.8[h]	60.8 ± 19.7[i]	60.6 ± 19.1[j]
ADCS-MCI-ADL-24	54.6 ± 8.2[f]	54.9 ± 7.9	54.5 ± 6.7[h]	55.4 ± 6.3	54.9 ± 7.3[j]
ADCOMS	0.4 ± 0.1[a]	0.4 ± 0.2[b]	0.4 ± 0.1[h]	0.4 ± 0.1[a]	0.4 ± 0.1[k]

Note: Data are mean ± standard deviation or n (%).

Abbreviations: ADAS-Cog14, Alzheimer's Disease Assessment Scale (14-Item) Cognitive Subscale; ADCOMS, Alzheimer's Disease Composite Score; ADCS-MCI-ADL-24, 24-item Alzheimer's Disease Cooperative Study/Activities of Daily Living Scale Adapted for Patients With Mild Cognitive 5 Impairment; CDR-SB, Clinical Dementia Rating-Sum of Boxes; FAQ, functional activities questionnaire; MMSE, Mini-Mental State Examination; NPI, Neuropsychiatric Inventory; RBANS, Repeatable Battery 7 for Assessment of Neuropsychological Status; TN, Tilavonemab; UPSA-Brief, University of California San Diego Performance-Based Skills Assessment-Brief 8 Version; y, years.

[a] $n = 75$.
[b] $n = 77$.
[c] $n = 73$.
[d] $n = 306$.
[e] $n = 76$.
[f] $n = 78$.
[g] $n = 310$.
[h] $n = 79$.
[i] $n = 74$.
[j] $n = 311$.
[k] $n = 306$.

parameters in Supplementary Table S1. Figure 1 (middle panels) display LOWESS curves for latent cognitive and functional severities from two separate unidimensional longitudinal IRT models, indicating a consistent decline across both severities.

3.4 | Multidimensional longitudinal IRT model for all CDR domains

Utilizing the multidimensional longitudinal IRT model for simultaneous evaluation of cognitive and functional severities, significant time effects were observed. Specifically, cognitive severity increased by 0.800 theta units annually (95% CI: [0.602, 1.006]) and functional severity by 0.620 theta units (95% CI: [0.433, 0.806]). This suggests cognitive decline surpasses functional impairment in early AD. Notably, tilavonemab doses did not significantly impact either severity, as indicated by 95% CIs of corresponding parameters including zero. Refer to Table 3 for regression parameters and Supplementary Tables S2 for location and discrimination parameters.

In line with unidimensional model findings, no significant correlation between baseline impairment and progression rate was noted in either severity, as evidenced by 95% CI for ρ_{01} and ρ_{23} including 0. Furthermore, a significant positive correlation between baseline cognitive and functional severities was identified ($\rho_{02} = 0.776$, 95% CI: [0.687, 0.850]), indicating synchronous baseline declines in both capabilities. A strong correlation was also found in their progression rates ($\rho_{13} = 0.964$, 95% CI: [0.908, 0.991]), suggesting aligned trajectories of deterioration. Detailed diagnostics for the multidimensional longitudinal IRT model are available in Supplementary Section 2, with Figures S4-S7 in the Supplemental Material underlining the model's robust fit and excellent performance in analyzing CDR domain scores. Figure 1 (lower panel) displays LOWESS curves for latent cognitive and functional severities from the multidimensional longitudinal IRT model, illustrating a consistent decline in both capabilities. Supplementary Figure S8

TABLE 2 Parameter estimates and 95% CIs from two separate unidimensional longitudinal IRT models for scores in cognitive and functional domains

Parameters	Mean	SD	95% CI	*P*	RSE
Cognitive model					
Time (y)	0.794	0.107	0.588, 1.004	<0.001	13.48
TN = 300 mg × time (y)	0.103	0.136	−0.152, 0.355	0.224	132.04
TN = 1000 mg × time (y)	0.036	0.136	−0.232, 0.301	0.395	377.78
TN = 2000 mg × time (y)	0.186	0.136	−0.082, 0.456	0.085	73.12
ρ	0.159	0.090	−0.015, 0.341	0.038	56.60
Functional model					
Time (y)	0.592	0.098	0.403, 0.796	<0.001	16.55
TN = 300 mg × time (y)	0.077	0.129	−0.166, 0.342	0.275	167.53
TN = 1000 mg × time (y)	0.180	0.128	−0.069, 0.427	0.080	71.11
TN = 2000 mg × time (y)	0.283	0.128	0.033, 0.542	0.014	45.22
ρ	−0.050	0.079	−0.205, 0.110	0.263	158

Note: ρ is the correlation coefficient of random intercepts and random slopes. *P*-value was calculated by: $p = 1 - \Phi(|\text{Mean}|/\text{SD})$, where $\Phi(\cdot)$ denotes the cumulative distribution function of standard Normal distribution. Refer to Model (2) in the Supplemental Materials for the meanings of β parameters.
Abbreviations: 95% CI, 95% credible intervals; *P*, *P*-value; RSE, relative standard error, computed as standard error divided by the absolute value of the estimate and multiplied by 100; SD, standard deviation; TN, tilavonemab; y, years;

TABLE 3 Parameter estimates and 95% CIs from the multidimensional longitudinal IRT model for scores in cognitive and functional domains

Parameters	Mean	SD	95% CI	*P*	RSE
Cognitive model					
Time (y)	0.800	0.104	0.602, 1.006	<0.001	13
TN = 300 mg × time (y)	0.087	0.139	−0.178, 0.362	0.265	159.77
TN = 1000 mg × time (y)	0.027	0.132	−0.238, 0.291	0.419	488.89
TN = 2000 mg × time (y)	0.182	0.136	−0.076, 0.454	0.090	74.73
ρ_{01}	0.143	0.080	−0.017, 0.298	0.037	55.94
σ_{u1}	0.486	0.086	0.337, 0.672	NA	17.70
Functional model					
Time (y)	0.620	0.094	0.433, 0.806	<0.001	15.16
TN = 300 mg × time (y)	0.040	0.126	−0.198, 0.295	0.375	315
TN = 1000 mg × time (y)	0.170	0.121	−0.060, 0.410	0.080	71.18
TN = 2000 mg × time (y)	0.220	0.125	−0.015, 0.467	0.039	56.82
ρ_{23}	−0.068	0.076	−0.210, 0.089	0.185	111.76
σ_{u3}	0.491	0.082	0.352, 0.669	NA	16.70
ρ_{02}	0.776	0.042	0.687, 0.850	<0.001	5.41
ρ_{03}	0.146	0.075	−0.003, 0.291	0.026	51.37
ρ_{12}	−0.017	0.077	−0.169, 0.129	0.413	452.94
ρ_{13}	0.964	0.021	0.908, 0.991	<0.001	2.18

Note: ρ_{01} is the correlation coefficient of random intercepts and random slopes in cognitive model; ρ_{02} is the correlation coefficient of random intercepts in cognitive model and random intercepts in functional model; ρ_{03} is the correlation coefficient of random intercepts in cognitive model and random slopes in functional model; ρ_{12} is the correlation coefficient of random slopes in cognitive model and random intercepts in functional model; ρ_{13} is the correlation coefficient of random slopes in cognitive model and random slopes in functional model; ρ_{23} is the correlation coefficient of random intercepts and random slopes in functional model; σ_{u1} is the standard error of the random slopes in cognitive model; σ_{u3} is the standard error of the random slopes in functional model. *P*-value was calculated by: $p = 1 - \Phi(|\text{Mean}|/\text{SE})$, where $\Phi(.)$ denotes the cumulative distribution function of standard Normal distribution. The *P*-values for σ_{u1} and σ_{u3} were not computed because the test statistic does not follow Normal distribution. Refer to Models (5) and (6) in the Supplemental Materials for the meanings of β parameters.
Abbreviations: 95% CI, 95% credible intervals; *P*, *P*-value; RSE, relative standard error, computed as standard error divided by the absolute value of the estimate and multiplied by 100; SD, standard deviation; TN, tilavonemab; y, years.

displays histograms of these severities, revealing no significant frequencies at the extremes, indicating minimum floor or ceiling effects. Additionally, Supplementary Table S3 presents the latent variables estimated by both unidimensional and multidimensional IRT models. The findings show that the estimated latent variables are consistent across the four treatment groups.

3.5 | Linear Mixed model on CDR-SB

Supplementary Table S4 presents the sensitivity analysis results of the linear mixed model on the CDR-SB scores. It suggests statistically significant time effect of deterioration in CDR-SB ($P<0.001$), but no significant effect from any tilavonemab dose. These results are consistent with our findings using both unidimensional and multidimensional longitudinal IRT models.

4 | DISCUSSION

In this research, we employed longitudinal IRT models to investigate the trajectories of cognitive and functional severities in early-stage AD patients and to assess the effects of tilavonemab. Our findings identified prominent temporal deteriorations in both cognitive and functional severities. Nevertheless, the therapeutic efficacy of tilavonemab in modulating these trajectories remained statistically insignificant. Notably, the initial level of impairments at baseline did not impact subsequent decline rates.

A salient observation from our analysis was the inherent association between cognitive and functional severities. Both severities exhibited similar patterns of impairment at the onset and followed consistent trajectories subsequently, highlighting the intercorrelated pathophysiology of cognitive and functional decline in AD. An important finding was the more rapid decline of cognition severity when compared to functional severity on the CDR in the early stages of AD. This parallels the clinical progression of disease, as outline in the FDA guidance, where inexorable cognitive decline occurs before functional impairments emerge. Recognizing these differential trajectories has significant clinical implications. A better understanding of accelerated cognitive declines can guide therapeutic strategies aimed at preserving cognitive functions, possibly improving the quality of life in early AD patients. In addition, in clinical trial designs, recognizing these trajectories can lead to improved patient stratification, which in turn can enhance the development and evaluation of therapeutic agents.

IRT modeling is a recognized methodological approach in AD research. Numerous studies have applied cross-sectional IRT models to evaluate scores from various assessments such as Alzheimer's Disease Assessment Scale-cognitive (ADAS-cog),[24,25] Alzheimer's Disease Cooperative Study-Activities of Daily Living inventory (ADCS-ADL),[26] and composite and other scales.[27–29] However, only four AD-centric publications have adopted longitudinal IRT models,[11,30–32] and all exclusively used the unidimensional variants. The multidimensional longitudinal IRT methodology, as applied herein, presents multiple advantages. It integrates diverse modeling components, offering a comprehensive insight into the temporal progression across the disease severity continuum. By accounting for inter-visit correlations within individual participants, it provides a detailed depiction of subtle changes across cognitive and functional severities. The methodology also permits concurrent estimation of specific domain-specific parameters and regression coefficients, yielding in-depth insights into domain attributes and determinants of severity progression. Moreover, the unitless attributes of the latent variables in the multidimensional model enhance the comparability of progression rates and tilavonemab effects across severities. To facilitate clinical interpretation of the regression coefficients expressed in unitless theta values, we can apply simulation techniques to express them in terms of CDR cognitive and functional subsums and total sum. These simulation techniques use the posterior samples from Bayesian inference and necessarily are associated with increased variance, as shown in our prior work.[17,18]

The evolution of AD is marked by distinct variabilities across individual patients and general heterogeneity within the broader patient population. Our model, which integrates random intercepts and slopes, effectively addresses these variabilities, offering individualized assessments of severity and progression across both cognitive and functional impairments. For a deeper insight into AD's progression, subsequent studies might expand the multidimensional longitudinal IRT model by incorporating inter-occasion random effects, thereby capturing the temporal variability inherent in AD progression. While our approach determined the number of latent variables anchored in the CDR definition and expert insights, an alternative is to employ an exploratory IRT model.[33] This model could autonomously define the optimal count of latent variables, facilitating a data-informed categorization of cognitive and functional impairments in AD's progression. Such a methodology promises to elucidate the intricate architecture and dynamics of AD, fostering a more profound comprehension and evaluation of its progression.

The AWARE study's restriction to Stage 3 early AD participants resulted in low counts in some categories, compelling us to consolidate scores into broader categories. This aggregation, while necessary for parameter estimation, may reduce the granularity and precision in evaluating the latent traits, and pose challenges in comparing our results with those from studies using different scoring methods. Future replication of our findings in larger datasets, ensuring adequate counts across all score categories, is desirable. Moreover, differential item functioning (DIF) arises in the CDR when item response probabilities of a domain vary among individuals with similar latent AD severity but differ by a covariate such as treatment group. While testing for DIF is crucial in scale validation, typically performed using cross-sectional IRT models as demonstrated in our previous studies,[34,35] current methodologies do not readily apply to the longitudinal IRT framework utilized in this research. This gap underscores the need for further methodological development to accurately assess DIF in longitudinal contexts. Last, to ensure model parsimony and computational efficiency, individuals with incomplete CDR data arrays were excluded from this analysis. While the Bayesian modeling framework is capa-

ble of handling missing data through posterior distribution sampling, this approach significantly increases the model's complexity and computational requirements. Future developments will aim to adapt our Bayesian modeling to comprehensively include missing data.

ACKNOWLEDGMENTS

Support for this work comes from National Institute on Aging (grants AG064803 and P30AG072958 to SL).

CONFLICT OF INTEREST STATEMENT

The authors have no potential conflicts of interest to report. Author disclosures are available in the supporting information.

CONSENT STATEMENT

Consent was not necessary for data analysis reported in this manuscript, as all data provided are deidentified and open source.

ORCID

Xiaoxiao Zhou https://orcid.org/0000-0003-3950-1980
Sheng Luo https://orcid.org/0000-0003-4214-5809

REFERENCES

1. Scheltens P, De Strooper B, Kivipelto M, et al. Alzheimer's disease. *Lancet*. 2021;397:1577-1590.
2. Morris JC. The Clinical Dementia Rating (CDR): current version and scoring rules. *Neurology*. 1993;43:2412-2414.
3. Lowe DA, Balsis S, Miller TM, Benge JF, Doody RS. Greater precision when measuring dementia severity: establishing item parameters for the Clinical Dementia Rating Scale. *Dement Geriatr Cogn Disord*. 2012;34:128-134.
4. Health UDo, Services H. Guidelines for Industry, Alzheimer's Disease: Developing Drugs for the Treatment of Early Stage Disease. 2013.
5. Health UDo, Services H. Early Alzheimer's Disease: Developing Drugs For Treatment, Guidelines for Industry. 2018.
6. Sevigny J, Chiao P, Bussière T, et al. The antibody aducanumab reduces Aβ plaques in Alzheimer's disease. *Nature*. 2016;537:50-56.
7. van Dyck CH, Swanson CJ, Aisen P, et al. Lecanemab in early Alzheimer's disease. *N Engl J Med*. 2023;388:9-21.
8. Florian H, Wang D, Arnold SE, et al. Tilavonemab in early Alzheimer's disease: results from a phase 2, randomized, double-blind study. *Brain*. 2023;146:2275-2284.
9. Jutten RJ, Papp KV, Hendrix S, et al. Why a clinical trial is as good as its outcome measure: a framework for the selection and use of cognitive outcome measures for clinical trials of Alzheimer's disease. *Alzheimers Dement*. 2023;19:708-720.
10. Dickson SP, Hogan NJ, Nicodemus-Johnson J, Knowlton N, Hennessey S, Hendrix SB. The CDR-SB and other clinical trial mistakes made due to regulatory momentum. *Alzheimers Dement*. 2022;18:e068296.
11. Vandemeulebroecke M, Bornkamp B, Krahnke T, Mielke J, Monsch A, Quarg P. A longitudinal item response theory model to characterize cognition over time in elderly subjects. *CPT Pharmacomet Syst Pharmacol*. 2017;6:635-641.
12. Bascoul-Mollevi C, Barbieri A, Bourgier C, et al. Longitudinal analysis of health-related quality of life in cancer clinical trials: methods and interpretation of results. *Qual Life Res*. 2021;30:91-103.
13. Li Y, Xiong C, Aschenbrenner AJ, et al. Item response theory analysis of the Clinical Dementia Rating. *Alzheimers Dement*. 2021;17:534-542.
14. Wang J, Luo S. Multidimensional latent trait linear mixed model: an application in clinical studies with multivariate longitudinal outcomes. *Stat Med*. 2017;36:3244-3256.
15. Wang J, Luo S, Li L. Dynamic prediction for multiple repeated measures and event time data: an application to Parkinson's disease. *Ann Appl Stat*. 2017;11:1787-1809.
16. Luo S, Zou H, Goetz CG, et al. Novel approach to movement disorder society-unified Parkinson's disease rating scale monitoring in clinical trials: longitudinal item response theory models. *Mov Disord Clin Pract*. 2021;8:1083-1091.
17. Luo S, Zou H, Stebbins GT, et al. Dissecting the domains of Parkinson's disease: insights from longitudinal item response theory modeling. *Mov Disord*. 2022;37:1904-1914.
18. Zou H, Aggarwal V, Stebbins GT, et al. Application of longitudinal item response theory models to modeling Parkinson's disease progression. *CPT Pharmacomet Syst Pharmacol*. 2022;11:1382-1392.
19. Zou H, Guo Y, Goetz CG, et al. Co-existent probable RBD and PD: disease progression, medication response, and clinical trial Implications. *Mov Disord Clin Pract*. 2023;11:312-314.
20. Harrell FE. *Regression modeling strategies: with applications to linear models, logistic regression, and survival analysis*. Springer; 2001.
21. Fox J-P. *Bayesian item response modeling: theory and applications*. Springer; 2010.
22. SD T. Stan modeling language. User's guide and reference manual. 2021.
23. Gelman A, Carlin JB, Stern HS, Rubin DB. *Bayesian data analysis*. Chapman and Hall/CRC; 1995.
24. Balsis S, Unger AA, Benge JF, Geraci L, Doody RS. Gaining precision on the Alzheimer's Disease Assessment Scale-cognitive: a comparison of item response theory-based scores and total scores. *Alzheimers Dement*. 2012;8:288-294.
25. Verma N, Markey MK. Item response analysis of Alzheimer's disease assessment scale. 2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE; 2014:2476-2479.
26. Ard MC, Galasko DR, Edland SD. Improved statistical power of Alzheimer clinical trials by item-response theory: proof of concept by application to the activities of daily living scale. *Alzheimer Dis Assoc Disord*. 2013;27:187.
27. Umeda-Kameyama Y, Mori T, Wada-Isoe K, et al. Development of a novel convenient Alzheimer's disease assessment scale, the ABC dementia scale, using item response theory. *Geriatr Gerontol Int* . 2019;19:18-23.
28. Snitz BE, Yu L, Crane PK, Chang C-CH, Hughes TF, Ganguli M. Subjective cognitive complaints of older adults at the population level: An item response theory analysis. *Alzheimer Dis Assoc Disord*. 2012;26:344.
29. Tsai C-F, Lee W-J, Wang S-J, Shia B-C, Nasreddine Z, Fuh J-L. Psychometrics of the Montreal Cognitive Assessment (MoCA) and its subscales: validation of the Taiwanese version of the MoCA and an item response theory analysis. *Int Psychogeriatr*. 2012;24:651-658.
30. 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;31:2152-2165.
31. Ueckert S. Modeling composite assessment data using item response theory. *CPT Pharmacomet Syst Pharmacol* . 2018;7:205-218.
32. Polhamus D, Dolton MJ, Rogers J, et al. Longitudinal exposure—response modeling of multiple indicators of Alzheimer's disease progression. *J Prev Alzheimers Dis* . 2023;10:212-222.
33. Fontanella L, Fontanella S, Valentini P, Trendafilov N. Simple structure detection through Bayesian exploratory multidimensional IRT models. *Multivar Behav Res*. 2019;54:100-112.
34. Goetz CG, Liu Y, Stebbins GT, et al. Gender-, age-, and race/ethnicity-based differential item functioning analysis of the movement disorder society-sponsored revision of the Unified Parkinson's disease rating scale. *Mov Disord*. 2016;31:1865-1873.

35. Luo S, Liu Y, Teresi JA, Stebbins GT, Goetz CG. Differential item functioning in the Unified Dyskinesia Rating Scale (UDysRS). *Mov Disord*. 2017;32:1244-1249.

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: Zhou X, Zou H, Lutz MW, et al. Assessing tilavonemab efficacy in early Alzheimer's disease via longitudinal item response theory modeling. *Alzheimer's Dement*. 2024;10:e12471. https://doi.org/10.1002/trc2.12471