J Prev Alz Dis 2023;2(10):212-222
Published online January 25, 2023, http://dx.doi.org/10.14283/jpad.2023.13

Original Research

Longitudinal Exposure–Response Modeling of Multiple Indicators of Alzheimer's Disease Progression

D.G. Polhamus[1], M.J. Dolton[2,], J.A. Rogers[1], L. Honigberg[2], J.Y. Jin[2], A. Quartino[2,**] for the Alzheimer's Disease Neuroimaging Initiative (ADNI)[***]*

1. Metrum Research Group, Tariffville, CT, USA; 2. Genentech, Inc., South San Francisco, CA, USA; *MJD was an employee of Genentech, Inc. at the time of this study (current affiliation: Roche Products Pty Limited, Sydney, Australia); **AQ was an employee of Genentech, Inc. at the time of this study (current affiliation: Clinical Pharmacology and Quantitative Pharmacology, AstraZeneca, Gothenburg, Sweden); ***Data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (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 list of ADNI investigators can be found at: http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf

Corresponding Author: Michael J. Dolton, Roche Products Pty Limited, Sydney, NSW, Australia; Telephone: +612 9454 9000; Email: michael.dolton@roche.com.

Abstract

BACKGROUND: Progression in Alzheimer's disease manifests as changes in multiple biomarker, cognitive, and functional endpoints. Disease progression modeling can be used to integrate these multiple measures into a synthesized metric of where a patient lies within the disease spectrum, allowing for a more dynamic measure over the range of the disease.

OBJECTIVES: This study aimed to combine modeling techniques from psychometric research (e.g., item response theory) and pharmacometrics (e.g., hierarchical models) to describe the multivariate longitudinal disease progression for patients with mild-to-moderate Alzheimer's disease. Additionally, we aimed to extend the subsequent model to make it suitable for clinical trial simulation, with the inclusion of covariates, to explain variability in latent progression (i.e., disease progression) and to aid in the assessment of enrichment strategies.

DESIGN: Multiple longitudinal endpoints in the Alzheimer's Disease Neuroimaging Initiative database were modeled. This model was validated internally using visual predictive checks, and externally by comparing data from the placebo arms of two Phase 2 crenezumab studies, ABBY (NCT01343966) and BLAZE (NCT01397578).

SETTING: The Alzheimer's Disease Neuroimaging Initiative began in 2004: the initial 5-year study (ADNI-1) was extended by 2 years in 2009 by a Grand Opportunities grant (ADNI-GO), and in 2011 and 2016 by further competitive renewals of the ADNI-1 grant (ADNI-2 and ADNI-3, respectively). This work studies natural progression data from patients with confirmed Alzheimer's disease. The Phase 2 ABBY and BLAZE trials evaluated the safety and efficacy of crenezumab in patients with mild-to-moderate Alzheimer's disease.

PARTICIPANTS: From the Alzheimer's Disease Neuroimaging Initiative database, 305 subjects who had a baseline diagnosis of mild-to-moderate Alzheimer's disease were included in modeling. From the ABBY and BLAZE studies, 158 patients were included from the studies' placebo arms.

MEASUREMENTS: Longitudinal cognitive and functional assessments modeled included the Clinical Dementia Rating (both as Sum of Boxes and individual item scores), the Mini-Mental State Examination, the Alzheimer's Disease Assessment Scale – Cognitive Subscale, the Functional Activities Questionnaire, the Montreal Cognitive Assessment, and the Rey Auditory Verbal Learning Test. Also included were the imaging variable fluorodeoxyglucose-positron emission tomography and the following magnetic resonance imaging volumetrics: entorhinal, fusiform, hippocampal, intra-cranial, mid-temporal, ventricular, and whole brain.

RESULTS: Applying item response theory approaches in this longitudinal setting showed clinical assessments informing a common disease scale in the following order (from early disease to late disease): Rey Auditory Verbal Learning Test, Functional Activities Questionnaire, Montreal Cognitive Assessment, Alzheimer's Disease Assessment Scale – Cognitive Subscale 12, Clinical Dementia Rating – Sum of Boxes, and Mini-Mental State Examination. The Clinical Dementia Rating communication and home-and-hobbies items were most informative at earlier disease stages, while memory, orientation, and personal care informed the disease status at later stages. A clinical trial simulation model was developed and accurately described within-sample longitudinal distribution of endpoints. Simplifying the model to use only baseline age, MMSE, and APOEε4 status as predictors, out-of-sample mean progression of ADAS-Cog and CDR Sum of Boxes in the ABBY and BLAZE placebo arms was accurately described; however, the variability in these endpoints was underpredicted and suggests possibility for further model refinement when extrapolating from the ADNI sample to trial data. Clinical trial simulations were performed to exemplify use of the model to investigate hypothetical disease modification effects on the multivariate, longitudinal progression on the Alzheimer's Disease Assessment Scale – Cognitive Subscale and the Clinical Dementia Rating – Sum of Boxes.

CONCLUSIONS: The latent variable structure of item response theory can be extended to capture a variety of scales that are common assessments and indicators of disease status in mild-to-moderate Alzheimer's disease. These models are not intended to support causal inferences, but they do successfully characterize the observed correlation between endpoints over time and result in concise numerical indices of disease status that reflect the totality of evidence from considering the endpoints jointly. As such, the models have utility for a variety of tasks in clinical trial design, including simulation of hypothetical drug effects, interpolation of missing data, and assessment of in-sample information.

Key words: Disease progression, item response theory, clinical trial simulation, prevention trials, Alzheimer's disease.

Received February 25, 2022
Accepted for publication October 16, 2022

Introduction

Changes in several endpoints (including cognitive, functional, and biomarker endpoints) are thought to be informative of disease progression in Alzheimer's disease (AD), but none in isolation are considered sufficient to define disease progression. Joint analysis of multiple endpoints may be more informative with regard to detecting and understanding disease progression. Given that the assessment of clinical trials in AD relies upon the accurate interpretation of multiple endpoints, as well as how they vary across the population of interest, the ability to understand, quantify, and simulate metrics of underlying disease status outcomes should enhance our ability to identify potentially successful therapies (i.e., identification of both disease progression and modification, which present across multiple endpoints) (1, 2).

There is a rich history of longitudinal modeling of univariate cognitive endpoints in AD. In the mild-to-moderate AD population, the Alzheimer's Disease Assessment Scale – Cognitive Subscale (ADAS-Cog) has historically served as the standard for assessment of cognition and has been described by multiple models that described versions of the summed score (ADAS-Cog 11) (3-5). The current trend toward the study of less-advanced disease populations has brought new focus on scales used for functional assessment and that are more sensitive in earlier disease states, such as the Clinical Dementia Rating (CDR) scale, which is typically analyzed as the "Sum of Boxes" (CDR-SB) (6-8), the Alzheimer's Disease Cooperative Study – Activities of Daily Living (ADCS-ADL), and the Functional Activities Questionnaire (FAQ) (9). Modeling the joint progression of these endpoints is a natural extension to these univariate approaches and is becoming an increasingly attractive method for understanding the long-term correlation between such endpoints, as target populations for potential treatments increasingly focus on individuals at earlier disease stages (10).

A psychometric item response theory (IRT) approach was applied in a novel longitudinal manner to determine sensitivity of the individual items within the ADAS-Cog score (11). IRT itself has been applied in non-longitudinal formats to CDR items to establish the ability of the different items to distinguish between patients at different disease stages, and also to demonstrate the similarity between this scale and another functional score (ADCS-ADL) (12).

The longitudinal IRT approach uses a latent structure to account for the dependence across items within a cognitive scale within a subject (11-13). A natural extension of this toward quantifying "disease progression" would be inclusion of multiple endpoints from different scales with different domains of data (e.g., a mixture of ordinal and continuous scales). A particular advantage to such a modeling approach is that the need for a rigorously structured dataset with simultaneous observation for multiple endpoints (as typically required in most multivariate regression scenarios) is not necessary (14). To accommodate this variety of data, Vandemeulebroecke and colleagues (15) as well as Leoutsakos and colleagues (16) applied the graded response model (a common IRT model) to a mixture of (naturally) ordinal data and ordinal transformations of continuous metrics of interest. Similar latent variable approaches have appeared in statistical literature in which discrete time latent variable models first described unbounded multivariate longitudinal data (17), and are now more frequently implanted in as joint models of longitudinal and time-to-event data (e.g., Tsiatis and Davidian (18)). A general longitudinal model for multivariate data of mixed types was introduced and compared to standard (fixed-time) IRT approaches (19).

We aimed to build upon previous work in longitudinal IRT approaches by preserving the original scales of the data, as would be desirable in the construction of clinical trial simulation (CTS) models. The objective was to develop a methodology that easily extended to new scales at earlier (or later) disease stages as understanding of the disease evolves over time, to support CTS across drug development programs. In developing this methodology, our goal was not primarily to judge the sensitivity of particular metrics of dementia (which is typically the goal of IRT), but to synthesize the information from the multiple metrics into a single latent measure of disease status while preserving the multivariate longitudinal correlation between the various metrics of the disease status. To this end, models were developed that included both endpoints commonly used to assess disease progression and drug effect in the clinical trial setting (e.g., CDR-SB, ADAS-Cog), but also progression markers thought to be more closely aligned with the disease mechanisms (e.g., fluorodeoxyglucose-positron emission tomography (FDG-PET) and magnetic resonance imaging (MRI)). Additionally, with the refined goal of building a model suitable for CTS in mind, a secondary modeling objective was to include covariates to explain variability in latent progression (i.e., disease progression) and to aid in assessment of enrichment strategies (3-5).

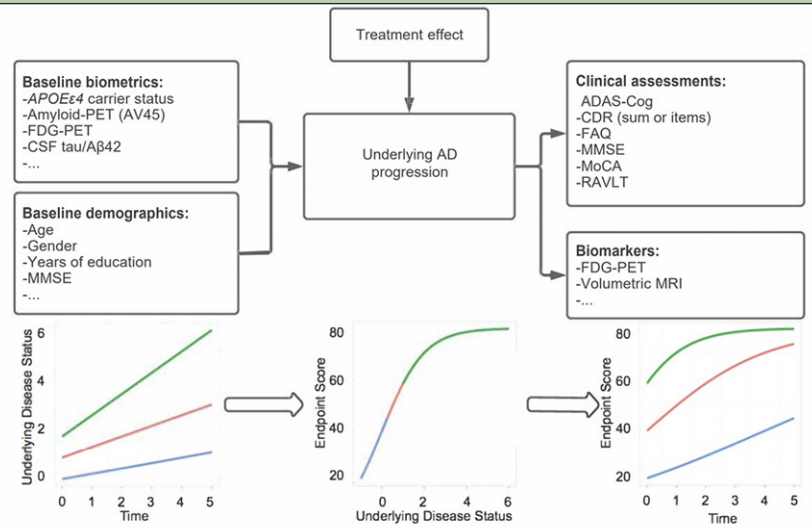
Methods

Data

Data used for the characterization of natural disease progression were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). ADNI is a longitudinal multicenter study designed to develop clinical, imaging, genetic, and biochemical biomarkers for the early detection and tracking of AD.

The ADNIMERGE R package (20) (Initiative 2014, packaged on January 20, 2014) was used to import data

Figure 1. Joint progression model



The general model architecture describes disease state as changing linearly over item, with intercepts and slopes adjusted by patient characteristics at study baseline. The linear progressions over time are then related to multiple endpoints that present as non-linear functions of the disease state and time. Disease-modifying therapy in this case is identified as a treatment that reduces the slope of the disease state progression over time. The bottom three panels illustrate the relationship between time, underlying disease status, and endpoint score for three patients at different disease states at baseline (time 0); Aβ, beta-amyloid; AD, Alzheimer's disease; ADAS-Cog, Alzheimer's Disease Assessment Scale – Cognitive Subscale; APOEε4, apolipoprotein Eε4 allele; CDR, Clinical Dementia Rating; CSF, cerebrospinal fluid; FAQ, Functional Activities Questionnaire; FDG-PET, fluorodeoxyglucose-positron emission tomography; MMSE, Mini-Mental State Examination; MoCA, Montreal Cognitive Assessment; MRI, magnetic resonance imaging; PET, positron emission tomography; RAVLT, Rey Auditory Verbal Learning Test.

into R (21) for both the initial 5-year study (ADNI-1) and a grant extension study (ADNI-2). The ADAS-Cog 12 score was derived by adding the delayed memory recall score (Q4) at each recording of the ADAS-Cog 11. Endpoints of interest selected for modeling were those used to assess efficacy in clinical trials in patients with prodromal and mild-to-moderate AD (mild cognitive impairment due to AD), as well as other measures of progression such as ADAS-Cog, CDR-SB, FAQ, the Mini-Mental State Examination (MMSE), the Montreal Cognitive Assessment (MoCA), and Rey Auditory Verbal Learning Test (RAVLT; immediate). In addition, longitudinal biomarkers of AD were included such as FDG-PET and volumetric MRI (MRI; entorhinal, fusiform, hippocampus, intracranial volume, mid-temporal, ventricular, and whole brain).

Data from the placebo arms of two Phase 2 crenezumab studies (ABBY [NCT01343966] and BLAZE [NCT01397578]) were used for external validation of the model. The only requirement to subject inclusion for either dataset was presence of at least one longitudinal (non-baseline) observation in any of the endpoints jointly modeled and, in the case of covariate inclusion, presence of the baseline covariate information. No explicit data imputation was performed.

Data assembly and post hoc analyses were performed using R 3.2.3 and model estimation was performed using OpenBUGS version 3.2.3 on an Ubuntu 12.04 LTS operating system. Details on computation can be found in the Appendix.

Model

The disease status was described using a latent process, following the conceptual approach used in IRT. As opposed to the typical longitudinal approach of modeling cognitive and functional endpoints as a function of time, the endpoints of interest were instead modeled as functions of the time-dependent latent disease status, which in turn can vary as a function of patient covariates, including treatment (Figure 1).

Patient latent disease progression was assumed to progress as a linear function of time since enrollment, with patient-specific intercepts to accommodate subjects entering a study at varying levels of the disease state (η_i; at study baseline) and slope (λ_i; see Ueckert et al. (11) and Polhamus et al. (22) for similar approaches). Let θi(t) represent the latent disease status for patient i at time t, then the progression over time is modeled as:

$$\theta_i(t) = \eta_i + \lambda_i t$$

As one of the aims of this modeling exercise is to simulate new trials with subjects who have a subset of the source data population characteristics, we model η_i and λ_i as a linear combination of covariates (x_i) and "fixed" effects (β).

Notably, the latent disease status for a reference patient (i.e., a hypothetical "typical" patient with reference covariate values) at baseline is defined to be standard normal for identifiability purposes. Having defined the latent progression over time, we then define submodels that describe the effect of the latent disease status on the

endpoints rather than on time, as done in more common longitudinal models. The Appendix describes the methodology used for the standard IRT graded-response model (i.e., ordered logit for ordinal endpoints, e.g., CDR items), a normal submodel (continuous on the real line), and a beta-residual submodel (for bounded responses, e.g., ADAS-Cog, MRI metrics, or CDR-SB).

Prior specification

A Bayesian probabilistic framework was established for estimation. The decision to use Bayesian methods was primarily motivated by a desire for the flexibility of estimation afforded by sampling methods, and only secondarily to allow the utilization of prior knowledge where applicable. Accordingly, the strategy for specification of priors can generally be classified as constraining the parameters to a subset that conveys plausible and meaningful clinical outcomes, yet remains otherwise non-informative (i.e., weakly informative). A notable exception here lies in the specification of the prior for $\beta((\lambda)$), the mean rate of progression on the latent scale, which was restricted to be positive for identifiability purposes. A list of the priors is available in the Appendix (Table A1).

Model building and validation

Two distinct modeling paths were pursued. The first pathway served to exemplify the ability of the model to capture the multivariate progression (the "joint progression model") of many endpoints measured at varying times and frequencies. As such, no covariate modeling was performed for this path; rather, the emphasis was on including all relevant longitudinal proxies of disease status with the largest patient subset as possible. In this path, all items were considered at the aggregate total level regardless of whether individual item scores were available (e.g., CDR-SB vs individual CDR items). Structural fit was assessed by comparing observed to predicted values conditional on the estimated random effects, including assessment of the preservation of correlation across endpoints by comparing the observed pairwise correlations (Spearman's rho) within individuals to those simulated from the model.

The second path (the "simulation model") aimed to assess the viability of the model for simulation purposes, by including common covariates used in clinical trial enrollment and stratification. Longitudinal endpoints included in this model reflected the available endpoints in the external validation dataset and are aligned with those metrics most commonly used as trial endpoints in mild cognitive impairment. The covariates selected reflect the availability in the external dataset, representing the well-known set of factors that characterize the rate of progression. These covariates included apolipoprotein E ε4 allele (APOEε4) carrier status, baseline MMSE (BMMSE), age, gender, years of education, baseline amyloid-PET (AV45), baseline FDG-PET, and baseline cerebrospinal fluid tau/beta-amyloid (Aβ)-42 ratios. Longitudinal endpoints included ADAS-Cog 12, the six individual CDR items, ventricular volume, and hippocampal volume. As individual CDR items are representative of different aspects of function and cognition, it was suggested that solitary items such as the memory subscore, in summation (CDR-SB) or in novel combinations, made for viable endpoints in a trial setting. The items were subsequently modeled individually using the graded-response submodel.

Model diagnostics were typical of mixed-effects regression models with emphasis on residuals (per endpoint) and random effects against observed values, time, and covariates of interest. Covariates were added to the model based upon visual inspection of trends in the residuals and random effects when compared to the covariates. Model comparison was driven by the deviance information criterion and inclusion of covariates thought to be relevant to trial conduct and simulation.

Model validation consisted of both internal and external simulation-based validation steps. For both validations, longitudinal statistics of interest were compared between the observed data and simulations from the model; coverage of the observed values by the 90% credible intervals (CrIs) of the simulations indicated model suitability. More detail can be found in the Appendix.

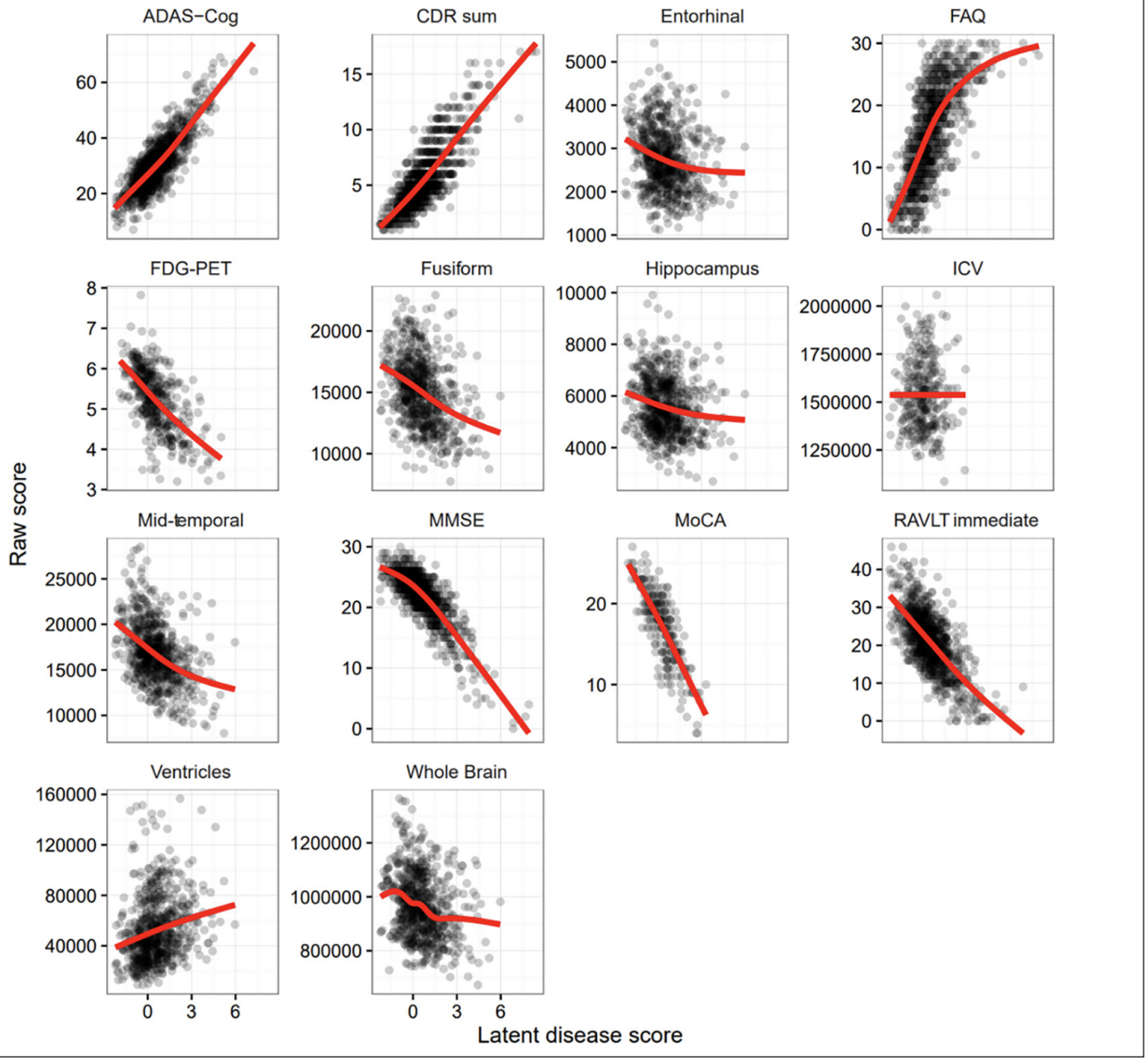
Information content

Information (statistical) was used to quantify the contribution of an endpoint to the latent scale, or regions of the latent scale. We used expected Fisher information for graded-response items as it is both analytically tractable and well known (23), but we used observed information for parameterizations lacking a simple analytical solution (i.e., the beta regression models). More details on each can be found in the Appendix.

Hypothetical drug effect

To assess the impact of a hypothetical disease-modifying drug, proportional reductions on the latent rate of progression were considered. In other words, $\theta\_i(t,\Delta)= \eta_i+ \lambda_i*(1-\Delta)*t$. Effects of size $\Delta= 0.2$, 0.4, and 0.6 (i.e., a 20%, 40%, and 60% disease-modifying reduction) were considered. To assess the effect in a real-world trial setting, the patients from the external validation set were used in 1000 simulated trials (i.e., covariates, dropout patterns, and follow-up duration from the external validation dataset were used as the virtual population). For each of the 1000 simulated trials, the mean change from baseline and mean percentage change from baseline were calculated and then summarized as medians and the 90% CrI at each nominal observation window.

Figure 2. The relationship between latent disease status and the endpoints used in the joint progression model



The ADNI joint progression model was fit to the set of endpoints shown below. The fitted value of the latent disease score was calculated at the posterior median for each observation and shown here. red lines are a loess smooth through the data; ADAS-Cog, Alzheimer's disease assessment scale-cognitive subscale; ADNI, Alzheimer's Disease Neuroimaging Initiative; CDR, Clinical Dementia Rating; FAQ, Functional Activities Questionnaire; FDG-PET, fluorodeoxyglucose-positron emission tomography; ICV, intracranial volume; MMSE, Mini-Mental State Examination; MoCA, Montreal Cognitive Assessment; RAVLT, Rey Auditory Verbal Learning Test.

Results

ADNI data for model building

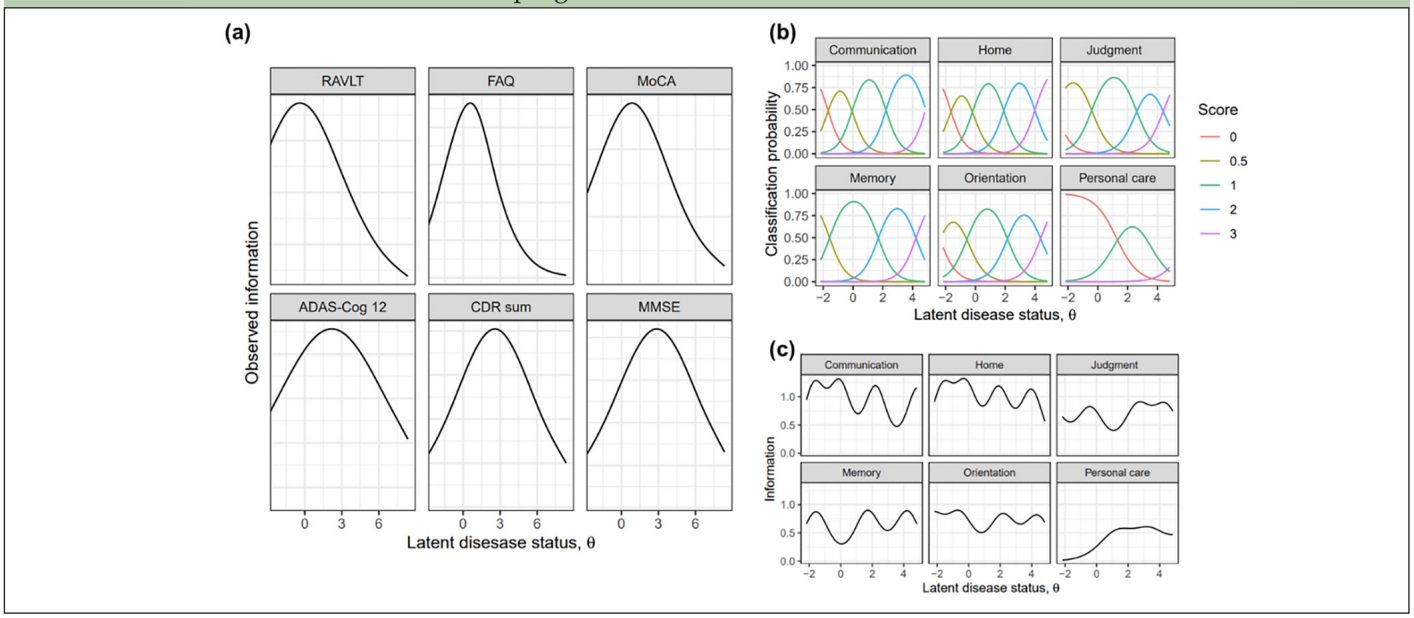
The ADNI database was subset of 305 subjects who had a baseline diagnosis of mild-to-moderate AD. ADAS-Cog mean at baseline was 28.0; most subjects (78%) had a CDR memory score of 1; 70% had at least one APOEε4 positive allele; other demographic summaries are shown in the Appendix (Table A2).

Joint progression model (Model 1)

Joint progression modeling was performed across 14 measures of mild-to-moderate AD, making no adjustment for covariates (as such, covariate effects, if present, are accounted for by the inter-individual random effects η_i, λ_i, and $\gamma(1,i)((l))$). Models with a mix of the beta-residual and normal residual submodels, and models with only beta-residual submodels were found to both provide similar fits. Model results using the beta-residual submodels are presented herein for a more direct comparison of the information functions (i.e., differences between the information curves are not due to model specification).

Posterior parameter estimates (Appendix, Table A3) of

Figure 3. IRT techniques were applied to examine the observed information (normalized) from the six clinical assessments included in the joint progression model (a) as well as the ICC (b) and FI (c) from the six CDR items as modeled in the simulation model. Disease progression increases with θ



ADAS-Cog 12, Alzheimer's Disease Assessment Scale – Cognitive Subscale 12; CDR, Clinical Dementia Rating; FAQ, Functional Activities Questionnaire; FI, Fisher information; ICC, item characteristic curve; IRT, item response theory; MMSE, Mini-Mental State Examination; MoCA, Montreal Cognitive Assessment; RAVLT, Rey Auditory Verbal Learning Test.

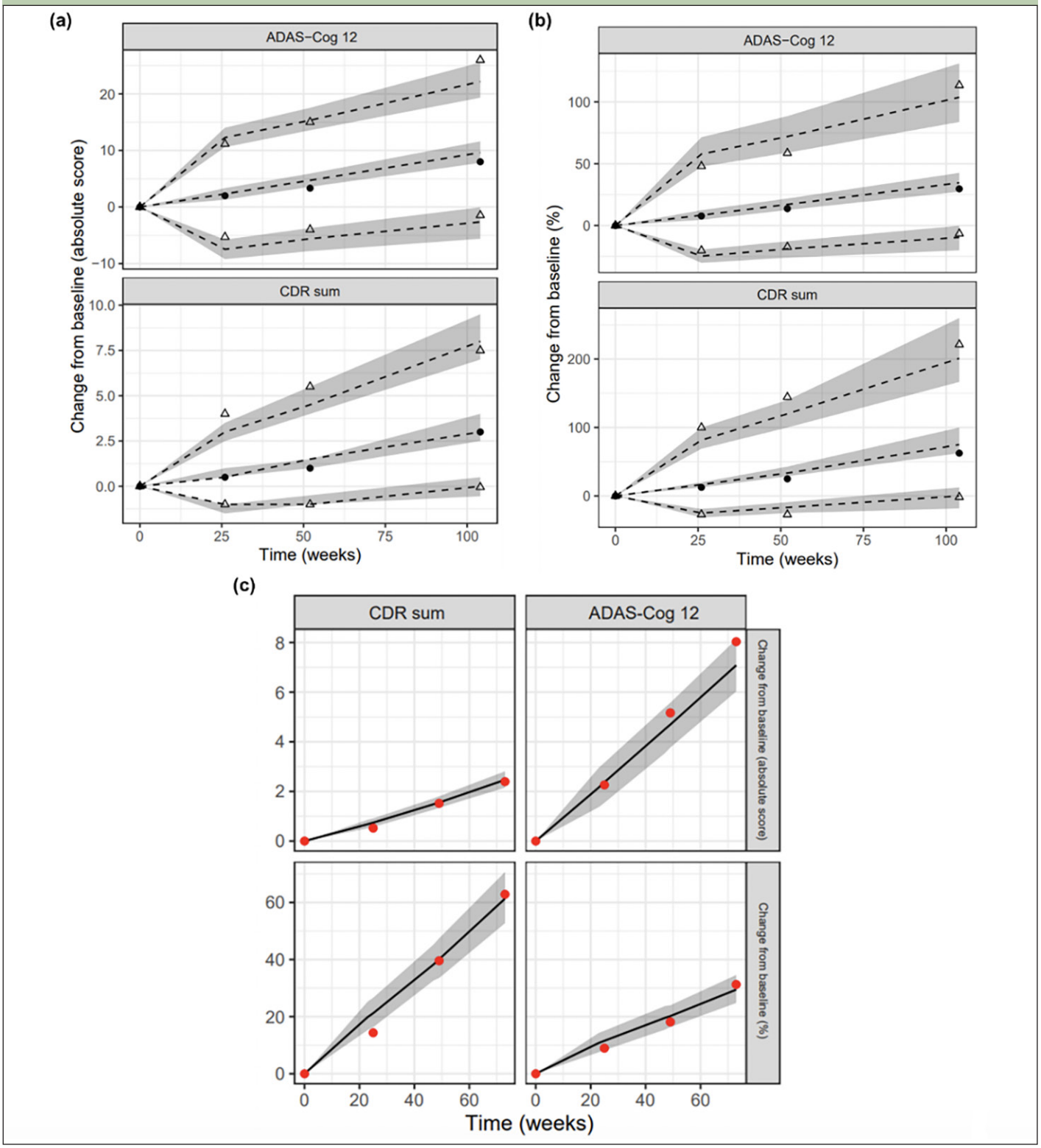
the endpoint-latent slopes (γ_2) for the joint-progression model fit increasingly positive values of the latent score to greater disease progression (Figure 2). Several of the cognitive and functional assessments show a strong correlation with the latent disease score, whereas the volumetric MRI results show more variability around — and less information on — latent disease (θ). This result is as expected, since cognitive and functional assessments, unlike volumetric MRI, have been developed specifically for the purpose of identifying disease in this population. A comparison of the observed and predicted data across all timepoints showed that the model captured the central tendency of the data well, and, as would therefore be expected, the multivariate correlation across endpoints was preserved. The normalized observed item information from each of the clinical assessments and cognitive tests are shown in Figure 3a. RAVLT, FAQ, and MoCA give higher amounts of information at earlier stages of disease as compared to ADAS-Cog, CDR, and MMSE.

Simulation model (Model 2)

For the second modeling objective, established markers and correlates of disease progression were considered on each of the η and λ structural parameters, including: APOEε4 carrier status, BMMSE, age, gender, years of education, baseline amyloid-PET (AV45), baseline FDG-PET, baseline cerebrospinal fluid tau/Aβ-42 ratios, and the two-way interactions between each of these covariates. FDG-PET in particular showed a high level of correlation with the latent slope parameter; however, due to the amount of missing data in the external validation dataset, only age, BMMSE, and APOEε4 status were retained in the final simulation model. All three of these factors affected disease progression, in that older APOEε4 positive females with a higher BMMSE progressed at a slower rate (see Appendix, Figure A1). However, due to their non-linear nature, the effect seen on these endpoints varied according to the disease state. For example, the effect of a one-unit difference in BMMSE is seen to have a larger difference for healthier (e.g., BMMSE ≥ 26) participants than those with a more advanced disease state (e.g., BMMSE ≤ 18) (Appendix, Figure A2). Volumetrics (hippocampal volume and ventricular volume) were included in the modeling of the ADNI data; however, they were found to be measured on a different scale in the external validation set and are therefore excluded from further discussion herein.

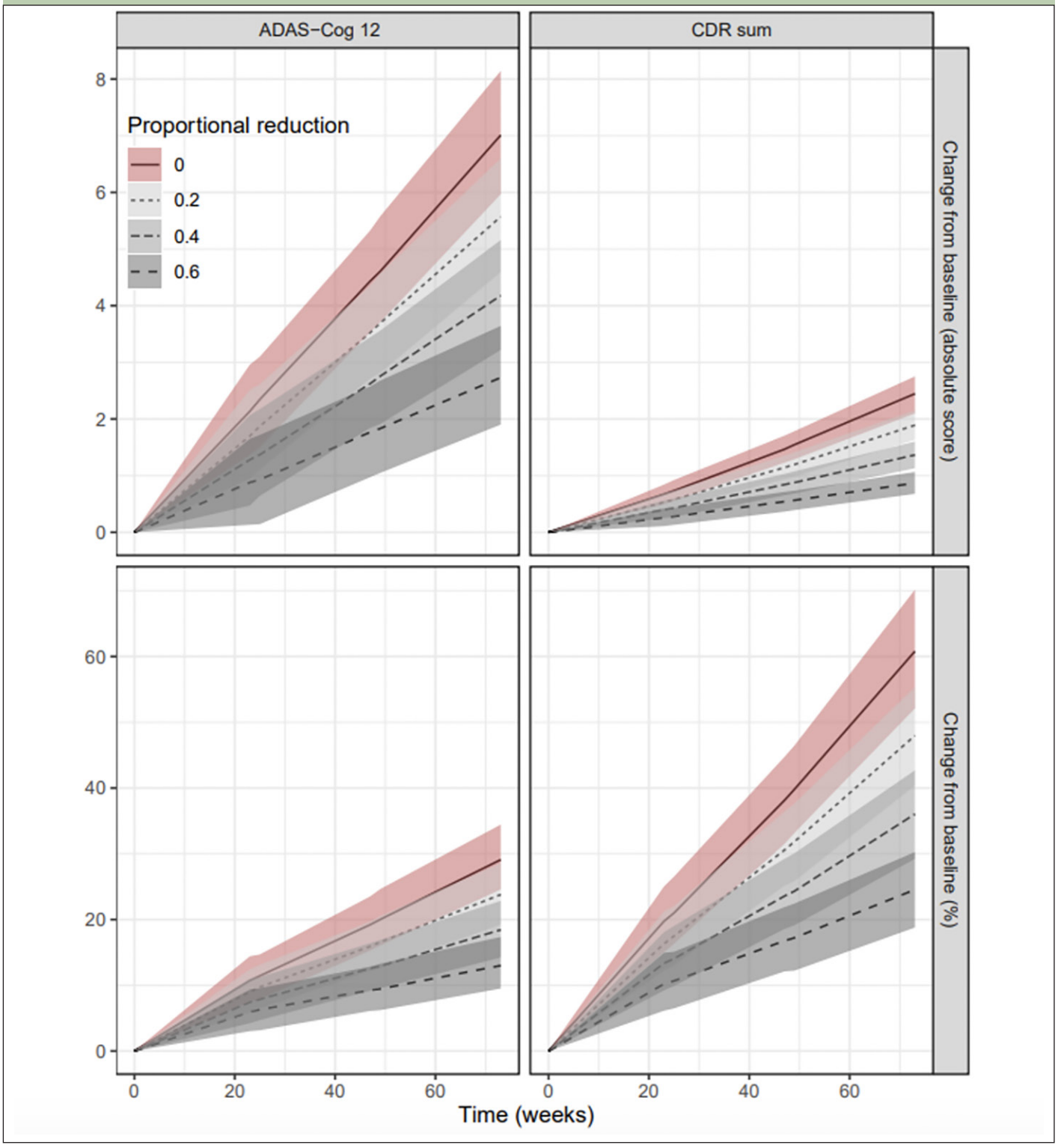
CDR was modeled with the graded-response model and the item characteristic curves and the expected Fisher information for each of the six items are shown in Figure 3b and c. As expected, the item characteristic curve plot shows that, with the ADNI mild-to-moderate AD subset, there are no patients with a memory score of 0 (the probability of getting that score is 0); as patients progress over the duration observed in this ADNI subset, 75% of the patients progress to a memory score of 3. Likewise, the judgment item has a low probability of patients having a score of 0. Most patients in the studied range of disease receive either a 0 or 1 score on the personal care item scale (i.e., this is not a sensitive measure within

Figure 4. Internal and external model validation for the simulation model. Internal visual predictive in which the median and 90% CrI around selected percentiles (the 5th, 50th, and 95th percentiles, circles and triangles for median and 5th, 95th respectively) for change from baseline (a) and percentage change from baseline (b) is compared with the observed quantities. Similarly, panel (c) shows an external validation using placebo arms from two Phase 2 trials



ADAS-Cog 12, Alzheimer's Disease Assessment Scale – Cognitive Subscale 12; CDR, Clinical dementia Rating; CrI, credible interval.

Figure 5. Change from baseline (absolute score) and change from baseline (%) in the presence of a disease-modifying drug effect. Drug effect as proportional reductions from 0.2 (or 20%) to 0.6 (60%) on the disease score slope, λ_i. The simulations were performed using the external validation set, i.e., patients from actual Phase 2 trials. The 5th, 50th, and 95th percentiles of the mean change are shown as the lines and shaded regions



ADAS-Cog 12, Alzheimer's Disease Assessment Scale – Cognitive Subscale 12; CDR, Clinical Dementia Rating.

this population), while the other items (community, home and hobbies, judgment, and orientation) show more sensitivity. The information curves corroborate these observations with communication and home-and-hobbies items both showing sensitivity across the studied range of conditions, but decreasing in disease severity. The judgment, memory, and orientation items all show lower information across the central area of the latent score distribution in this sample. Personal care has little information on θ until the latest stages studied.

Model validation

Model validation was performed using the simulation model (Model 2). For the purpose of internal validation, visual predictive checks (VPCs) were created for absolute values, change from baseline, and percentage change from baseline values. Additionally, VPCs by covariate strata were generated. All VPCs suggest that the model was able to replicate the data and is fit for purpose for simulation (Figure 4). For external validation, 158 patients with mild-to-moderate AD were used from the Phase 2 studies (Appendix, Table A4). Compared with the ADNI patient population, the Phase 2 patient population was numerically younger than the ADNI population (mean and range of 70.1 [51-80] for crenezumab vs 74.9 [55.1-90.9] for ADNI) and was more evenly balanced across gender (53.8% female for crenezumab vs 44.6% female for ADNI), but were otherwise similar. Results of the external validation are shown in Figure 4. The simulated endpoints were assessed for the mean of the change from baseline scores as well as the percentage change from baseline scores. CDR was summarized as CDR-SB rather than the individual item scores, to more closely resemble a trial efficacy endpoint. Regarding the central tendency, the model predicted more progression in CDR-SB at week 20 than what was observed, and likewise under-predicted the week 64 ADAS-Cog change from baseline (although not the percentage change from baseline).

Hypothetical drug effect

Results from the simulation of a disease-modifying drug's effect are shown in Figure 5. A 20% reduction showed overlapping 90% CrIs with placebo for both ADAS-Cog and CDR-SB. At week 73, the placebo progression for ADAS-Cog percentage change from baseline was 29.1% (90% CrI 24.6%, 34.4%) as compared to a 23.8% (90% CrI 19.3%, 28.5%) change under the hypothetical 20% reduction. These correspond to change from baseline scores of 7.0 (90% CrI 6.0, 8.1) points for no disease modification and 5.6 (90% CrI 4.6, 6.6) points for a 20% reduction. For CDR-SB, the change from baseline for placebo patients was simulated as 2.4 (90% CrI 2.1, 2.8) or 60.8% (90% CrI 52.1%, 70.2%) as compared to a 1.9 (90% CrI 1.6, 2.2) or 48.0% (90% CrI 40.4%, 55.3%) change with a disease-modifying effect of 20%. At a 40% reduction of disease progression, the endpoints were significantly different based upon the 90% CrIs.

Discussion

Extending the longitudinal latent variable model framework used in IRT can successfully capture both within-sample and out-of-sample progression of multiple endpoints of interest in mild-to-moderate AD. The out-of-sample prediction had good coverage of the central tendency but under-predicted the variability in the external sample: a possible explanation for this could be that the ADNI data were not generated under a trial setting and instead reflect natural disease progression. Additionally, the data were subject to a different set of entry criteria and trial logistics that plausibly were not captured by the limited set of covariates included in modeling. The benefits of modeling the multivariate progression (as opposed to using univariate models) are many, and include assessment of information content from endpoints on the progression of a population of interest, a cohesive methodology to synthesize the information from multiple endpoints with varying missingness patterns (missing data do not bias our inferences here, under the assumption of a "Missing at Random" missing data mechanism (24)) into a single measure of disease status, and the ability to then simulate multivariate patient data that preserve the correlation in the endpoints across endpoints and time. The summary of progression as a longitudinal univariate score allows for simple assessment of disease progression across multiple endpoints. For example, using CTSs we found that the detection of a 20% reduction in disease progression in a balanced study of patients with mild-to-moderate AD (at 158 patients per arm) was unlikely based upon ADAS-Cog and the CDR-SB.

Interventions modeled as acting directly upon the latent scale should be considered with care, as this induces an averaging of the effect across the modeled endpoints. When including endpoints with varying degrees of sensitivity to the intervention, fitted values of the more sensitive endpoints (to the intervention) will tend to show under-prediction of the effect, while the less-sensitive measures will show a bias towards over-prediction. Reducing these biases in the effects of the intervention may be preferable in situations such as CTSs and can be addressed by modeling the effect separately per endpoint (i.e., within the submodels), with subsequent assessment of differing levels of disease modification by endpoint.

When jointly modeling all of the endpoints relevant to this patient population in the joint progression model, many of the information curves correspond to our current understanding of where these scales fit within the disease spectrum. While the CDR-SB is typically thought to be most informative for the prodromal population, it appears here to provide information on a similar disease

stage to that of ADAS-Cog and MMSE. Extrapolating the observed information plots to earlier disease states, one would expect RAVLT, FAQ, and MoCA to be increasingly useful to detect progression in patients at early stages with ADAS-Cog 12, CDR-SB, and MMSE being more discriminatory at later stages.

As clinical trials are conducted in patient populations with increasingly milder forms of AD, the required duration of a trial to observe changes in rates of progression per treatment will increase. In longer trials, a metric that began a study as being most informative may not remain the best marker of progression by the trial's end. In these increasingly common situations, models that synthesize information onto a univariate scale could offer a way to detect and track the continual changes in progression. Coupled with dose/exposure–response modeling, the assessment of how early or late a drug becomes efficacious may help to optimize treatment across the different stages of the disease. In these longer trial settings, and despite not being evident in this work, it is also possible that the assumption of linear progression of the latent variable in time is overly simplistic and that such progression may be more apparent over longer time intervals and longer predictions. An adaptation of the model to utilize a latent variable process and allow for time non-linearity (25) would be a good direction of future work; however, the challenge of identifiability between non-linearity of the scales and non-linearity in the latent variable seems a likely problem.

Modeling herein was performed in a Bayesian context to leverage the flexibility of sampling-based estimation. Doing so requires specification of priors; prior selection here was made in the context of internal decision-making, with simulation-based diagnostics used to ascertain adequacy of the model for simulation purposes, which was the primary goal of this analysis. More extensive sensitivity analyses regarding the priors, including use of prior-predictive simulations, would be necessary in order to ensure fitness for purpose with a broader set of stakeholders.

We have demonstrated that this modeling method gives a model that is able to replicate the observed multivariate trajectories. A possible beneficial use of such a multivariate model would be in the framework of multivariate decision or testing processes, such as investigating the success rate for a planned gatekeeping procedure based upon multiple endpoints. Extending the longitudinal latent variable approach to allow for simulation of doses, patient populations, and trial parameters outside of those studied while preserving this correlation leads to more accurate assessment of the performance of such testing procedures.

Funding: 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 the DOD ADNI (Department of Defense award number W81XWH-12-2-0012). The Metrum Research Group (Tariffville, CT, USA) and Genentech, Inc. (South San Francisco, CA, USA) were involved in the funding of the work, and the analysis and interpretation of the data. The ABBY and BLAZE studies were funded by Genentech, Inc. Editorial support in the preparation of this manuscript was funded by F. Hoffmann-La Roche Ltd, Basel, Switzerland.

Acknowledgments: We would like to thank all Alzheimer's Disease Neuroimaging Initiative (ADNI), ABBY, and BLAZE participants, clinicians, and administrators. Data collection and sharing for this project was funded by the ADNI (National Institutes of Health Grant U01 AG024904) and the DOD ADNI (Department of Defense award number W81XWH-12-2-0012). The 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 provides 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). 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 of Neuro Imaging at the University of Southern California. Crenezumab was discovered and developed in collaboration with AC Immune SA, Lausanne, Switzerland. Editorial support in the development of this manuscript was provided by Chris Ackroyd, MSc, of Health Interactions, funded by F. Hoffmann-La Roche Ltd, Basel, Switzerland.

Conflict of interest: DGP and JAR are employees of the Metrum Research group. MJD was an employee of Genentech, Inc., a member of the Roche Group, at the time of this work; he is now an employee of Roche Products Pty Limited, Sydney, Australia, and owns stock or stock options in F. Hoffmann-La Roche Ltd. LAH and JYJ are employees of Genentech and own stock in F. Hoffmann-La Roche Ltd. AQ was an employee of Genentech, Inc. and owned stock in Roche Holding Ltd at the time of this study; she is now an employee of AstraZeneca, Gothenburg, Sweden.

Data availability statement: For up-to-date details on Roche's Global Policy on the Sharing of Clinical Information and how to request access to related clinical study documents, see here: https://go.roche.com/data_sharing. Requests for the modeling data underlying this publication requires a detailed, hypothesis-driven statistical analysis plan that is collaboratively developed by the requestor and company subject matter experts. Direct such requests to global.patient_level_data_sharing@roche.com for consideration. Anonymized records for individual patients across more than one data source external to Roche can not, and should not, be linked due to a potential increase in risk of patient re-identification. Data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (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 list of ADNI investigators can be found at: http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf

Ethical standards: All human procedures were conducted in accordance with the Declaration of Helsinki and Good Clinical Practice.

References

1. Sun B-L, Li W-W, Zhu C, et al. Clinical research on Alzheimer's disease: progress and perspectives. Neurosci Bull 2018;34:1111-1118.
2. Hamasaki T, Evans SR, Asakura K. Design, data monitoring, and analysis of clinical trials with co-primary endpoints: a review. J Biopharm Stat 2018;28:28-51.
3. Samtani M. Disease progression analysis for ADNI MCI subjects utilizing ADAS-cog/11 scores. Creating Consensus Science: New Tools and Tactics for Next-Gen Drug Development. Session IV. Critical Path Institute. 2011. p31–41: https://c-path.org//wp-content/uploads/2013/09/consensussciencequantitativediseaseprogressionmodelsastools.pdf. Accessed 15 July 2022.
4. Rogers JA, Polhamus D, Gillespie WR, et al. Combining patient-level and summary-level data for Alzheimer's disease modeling and simulation: a β regression meta-analysis. J Pharmacokinet Pharmacodyn 2012;39:479-498.
5. Ito K, Ahadieh S, Corrigan B, et al. Disease progression meta-analysis model in Alzheimer's disease. Alzheimers Dement 2010;6:39-53.
6. Delor I, Charoin JE, Gieschke R, Retout S, Jacqmin P. Modeling Alzheimer's disease progression using disease onset time and disease trajectory concepts applied to CDR-SOB scores from ADNI. CPT Pharmacometrics Syst Pharmacol 2013;2:e78.

7. Morris JC. The Clinical Dementia Rating (CDR): current version and scoring rules. Neurology 1993;43:2412-2414.
8. O'Bryant SE, Lacritz LH, Hall J, et al. Validation of the new interpretive guidelines for the clinical dementia rating scale sum of boxes score in the national Alzheimer's coordinating center database. Arch Neurol 2010;67:746-749.
9. Pfeffer RI, Kurosaki TT, Harrah CH Jr, Chance JM, Filos S. Measurement of functional activities in older adults in the community. J Gerontol 1982;37:323-329.
10. Li D, Iddi S, Thompson WK, et al. Bayesian latent time joint mixed-effects model of progression in the Alzheimer's Disease Neuroimaging Initiative. Alzheimers Dement (Amst) 2018;10:657-668.
11. 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.
12. Miller TM, Balsis S, Lowe DA, Benge JF, Doody RS. Item response theory reveals variability of functional impairment within clinical dementia rating scale stages. Dement Geriatr Cogn Disord 2011;32:362-366.
13. 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.
14. Rencher AC, Christensen WF. Methods of multivariate analysis. 2012. Wiley, Hoboken, New Jersey.
15. Vandemeulebroecke M, Bornkamp B, Krahnke T, et al. A longitudinal item response theory model to characterize cognition over time in elderly subjects. CPT Pharmacometrics Syst Pharmacol 2017;6:635-641.
16. Leoutsakos JM, Gross AL, Jones RN, Albert MS, Breitner JCS. 'Alzheimer's Progression Score': development of a biomarker summary outcome for AD prevention trials. J Prev Alzheimers Dis 2016;3:229-235.
17. Roy J, Lin X. Latent variable models for longitudinal data with multiple continuous outcomes. Biometrics 2000;56:1047-1054.
18. Tsiatis A, Davidian M. Joint modeling of longitudinal and time-to-event data: an overview. Stat Sin 2004;14:809-834.
19. Proust-Lima C, Amieva H, Jacqmin-Gadda H. Analysis of Multivariate Mixed Longitudinal Data: A Flexible Latent Process Approach. Br J Math Stat Psychol 2013;66:470-487.
20. Alzheimer's Disease Neuroimaging Initiative (ADNI). ADNI Data Package for R. ADNIMERGE 0.0.1. 2021. https://adni.bitbucket.io. Accessed 15 July 2022.
21. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2020. https://www.R-project.org/. Accessed 15 July 2022.
22. Polhamus DG, Rogers JA, Gillespie WR, French J. Clinical dementia rating modeling and simulation: joint progression of CDR and biomarkers in the ADNI cohort. 2013. www.metrumrg.com/wp-content/uploads/2017/10/Polhamus_AAIC_2013.pdf. Accessed 15 July 2022.
23. van der Linden WJ, Pashley PJ. Elements of Adaptive Testing. In: van der Linden WJ, Glas CAW (eds) Item Selection and Ability Estimation in Adaptive Testing. 2009. Springer New York, New York, NY, pp 3-30.
24. Gastonguay MR, French JL, Heitjan DF, et al. Missing data in model-based pharmacometric applications: points to consider. J Clin Pharmacol 2010;20:63S-74S.
25. Proust-Lima C, Dartigues J-F, Jacqmin-Gadda H. Joint modeling of repeated multivariate cognitive measures and competing risks of dementia and death: a latent process and latent class approach. Stat Med 2016;35:382-398.
26. Sevcikova H, Rossini T, L'Ecuyer P. Package 'rlecuyer'. 2019. https://cran.r-project.org/web/packages/rlecuyer/rlecuyer.pdf. Accessed 15 July 2022.

How to cite this article: D.G. Polhamus, M.J. Dolton, J.A. Rogers, et al. Longitudinal Exposure–Response Modeling of Multiple Indicators of Alzheimer's Disease Progression. J Prev Alz Dis 2023;2(10):212-222; http://dx.doi.org/10.14283/jpad.2023.13