Received: 20 March 2020 | Revised: 18 December 2020 | Accepted: 15 February 2021

DOI: 10.1002/sim.8932

RESEARCH ARTICLE

Statistics in Medicine WILEY

Simultaneous modeling of Alzheimer's disease progression via multiple cognitive scales

Line Kühnel[1,2] | **Anna-Karin Berger**[1] | **Bo Markussen**[2] | **Lars L. Raket**[1,3]

[1]H. Lundbeck A/S, Valby, Denmark

[2]Department of Mathematical Sciences, University of Copenhagen, Copenhagen, Denmark

[3]Clinical Memory Research Unit, Department of Clinical Sciences, Lund University, Lund, Sweden

Correspondence
Line Kühnel, H. Lundbeck A/S, Valby, Denmark.
Email: LIHK@lundbeck.com

Lars L. Raket, H. Lundbeck A/S, Valby, Denmark.
Email: stat@larslau.dk

Funding information
Innovationsfonden, Grant/Award Number: 9066-00005B

Analyzing the progression of Alzheimer's disease (AD) is challenging due to lacking sensitivity in currently available measures. AD stages are typically defined based on cognitive cut-offs, but this results in heterogeneous patient groups. More accurate modeling of the continuous progression of the disease would enable more accurate patient prognosis. To address these issues, we propose a new multivariate continuous-time disease progression (MCDP) model. The model is formulated as a nonlinear mixed-effects model that aligns patients based on their predicted disease progression along a continuous latent disease timeline. The model is evaluated using long-term follow-up data from 2152 participants in the Alzheimer's Disease Neuroimaging Initiative. The MCDP model was used to simultaneously model three cognitive scales; the Alzheimer's Disease Assessment Scale-cognitive subscale, the Mini-Mental State Examination, and the Clinical Dementia Rating scale—sum of boxes. Compared with univariate modeling and previously proposed multivariate disease progression models, the MCDP model showed superior ability to predict future patient trajectories. Finally, based on the multivariate disease timeline estimated using the MCDP model, the sensitivity of the individual items of the cognitive scales along the different stages of disease was analyzed. The analysis showed that delayed memory recall items had the highest sensitivity in the early stages of disease, whereas language and attention items were sensitive later in disease.

KEYWORDS

Alzheimer's disease, cognitive assessment, disease progression model, item analysis, multivariate analysis, nonlinear mixed-effects model, ordinal model

1 | INTRODUCTION

Alzheimer's disease (AD) is a slowly progressing disease where affected individuals progress from having normal cognition to severe dementia over a period that is typically more than a decade.[1] However, lack of sensitivity of currently available measures and considerable individual differences in disease manifestation make it difficult to accurately stage patients and predict their future course of disease. The progression of AD is typically evaluated by repeated assessments of clinical scales measuring cognition or function, for example, the Alzheimer's Disease Assessment Scale-cognitive subscale (ADAS-cog).[2] However, there has recently been an increasing focus on biomarkers for measuring disease progression, with a specific focus on measures of amyloid and tau burden and neurodegeneration biomarkers.[3,4] While a number of

these biomarkers have been shown to be of great diagnostic value, their prognostic value compared with clinical scales is yet to be thoroughly investigated.

The current inability to define patient populations with homogeneous progression patterns makes it particularly difficult to conduct interventional clinical trials in AD. A heterogeneous trial population will have considerable variation in the observed patterns of decline due to disease-stage differences, which will in turn reduce power to detect treatment effects. This is illustrated by a recent power calculation suggesting that to obtain a reasonable power for detecting a treatment effect on cognitive performance in preclinical AD, clinical trials should run for at least 4 years, with a sample size of approximately 2000 patients per arm.[5] For drug development sponsors, such trials would typically not be considered economically viable. However, there are at least three widely investigated ways to decrease trial durations while retaining power[6-11]: (1) Use inclusion criteria that reduce heterogeneity of the included patient populations; (2) Develop more sensitive measures of progression in the early stages of disease; and (3) Identify subpopulations with accelerated disease progression.

While these strategies have been widely investigated, the analyses often rely on grouping patients into coarse disease-stage groups (eg, mild cognitive impairment, mild/moderate/severe AD dementia) and oversimplistic statistical models. For example, a commonly used procedure for identifying fast-progressing patients is to predict rate of change on cognitive measures using random slope models. The patients with the steepest predicted slopes are identified as the fast progressors.[12,13] However, because the rate of cognitive decline is typically increasing as AD progresses, this approach tends to identify patients that are late in the disease. Even, when adjusting analyses for baseline cognition, this bias persists since cognitive scores are not only affected by disease stage but also by factors such as cognitive reserve.[14] Consequently, the patients that are later in disease (with an associated faster rate of decline) but better able to compensate on cognitive tests will look like the fastest progressors. This may explain why longer education and occupational complexity have consistently been found to be associated with an increased rate of cognitive decline in AD.[15]

There exist a range of models for analyzing disease progression in AD. Continuous-time linear mixed-effects models provide a simple framework for modeling progression, but they will generally not enable separation of (horizontal) disease-stage effects from (vertical) additive effects since everything is modeled on an additive scale. The mixed-model for repeated measures[16] treats time as a categorical variable which enables the approximation of any nonlinear time pattern. However, since time is considered categorical it is difficult to model horizontal disease-stage effects, as any systematic variation from the response is modeled as a vertical additive effect.

An early example of explicitly modeling patient-level disease stage is that provided by Jedynak et al[17] Donohue et al[18] developed a similar modeling approach, the so-called growth models by alternating conditional expectation (GRACE) framework. More recently, Li et al[19] developed a class of latent-time joint mixed-effects models (LTJMM) for modeling nonlinear disease progression on multiple outcomes. In LTJMM, disease stage differences are modeled by a subject-specific random time shift. The nonlinear progression curve is achieved by nonlinear transformation of outcomes prior to analysis, and potentially the use of a link function. These transformed outcomes are modeled using a linear mixed-effects model and the identifiability of the latent time shifts is achieved by assuming that disease stage is the only consistent time-invariant patient-level effect across outcomes. Another recent disease progression modeling framework has been proposed by Raket.[20] This framework considers the same problem but differ in some key modeling choices. Most notably, the model uses nontransformed data and directly models the nonlinear disease progression curve that relates the observed outcomes on their original scale to the latent disease time. This enables flexible and interpretable noise models to be specified and allow separation of time-invariant effects of disease stage from other additive effects. Furthermore, the framework allows modeling of covariate effects on both disease stage, rate of decline and deviation from the mean. However, the models proposed by Raket do not allow simultaneous modeling of multiple outcomes.

This article extends the univariate framework proposed by Raket to a new multivariate continuous-time disease progression (MCDP) model that can handle multivariate outcomes on their original scales. This enables the use of information from several cognitive measures that may differ in sensitivity across disease stages. The multivariate analyses presented here are based on three cognitive measures ADAS-cog, Mini-Mental State Examination (MMSE) [10], and the Clinical Dementia Rating scale—sum of boxes (CDR-SB).[11] The MCDP model was fitted to data from the Alzheimer's Disease Neuroimaging Initiative (ADNI) and compared with corresponding univariate models to investigate the value of simultaneous using information from multiple cognitive measures. The MCDP model was subsequently compared with multivariate modeling results obtained from GRACE and LTJMM in both a prediction study based on ADNI data and in a simulation study. Finally, based on the predicted patient staging results from the MCDP model, the temporal sensitivity patterns of items from the three cognitive measures were compared along the common continuous disease timeline. This

common timescale was used to estimate the continuous-time evolution of each individual item enabling comparison of item sensitivity across different cognitive assessments.

The rest of the article is structured as follows. Section 2 includes a description of the analyzed data; a presentation of the modeling framework; and a presentation of the methodology used for validating and comparing model results. The results of the multivariate analysis of cognitive measures and item sensitivities are described in Section 3, followed by a discussion of the results in Section 4.

2 | METHODS

2.1 | Data

Data from the ADNI database (adni.loni.usc.edu) were used for the analyses presented in this article. The database holds data from patients with varying degrees of cognitive impairment, with follow-up times of more than 10 years. The analyses are based on three cognitive scales, ADAS-cog, MMSE, and CDR-SB. ADAS-cog ranges from 0 to 85 points with higher values indicating more severe impairment. The MMSE score ranges from 0 to 30 points with lower values implying more severe impairment. The CDR-SB score ranges from 0 to 18 points with 0 indicating no dementia and 18 severe dementia.

To be included in the analysis, patients were required to have a valid disease severity status at baseline in one of the five categories; cognitively normal, significant memory concern, mild cognitively impaired (early), mild cognitively impaired (late), or dementia. Furthermore, included patients were required to have had at least one measurement of one of the three cognitive scales. These inclusion criteria were fulfilled by 2152 of 2175 participants. Among the 31 excluded participants, 23 had a missing baseline disease severity status and the remaining eight did not have any valid observations of the cognitive scales. The 2152 included participants had a total of 10 160 visits with valid cognitive assessments. All models were fitted on a randomly selected subset of 80% of participants (1722 participants, 8044 visits). The remaining 20% of participants (430 participants, 2116 visits) were held out for validation.

2.2 | Disease progression modeling

Assume that subject i has m_i visits at time points $t_{i1} < \ldots < t_{im_i}$. Both number and timing of visits can vary across subjects. Let y_{ij} denote the observed outcome from subject i at time t_{ij}.

We consider a univariate disease progression model of the form

$$y_{ij} = \mu(t_{ij} + \boldsymbol{x}_{ij}^T\beta + z_i) + v_i + \varepsilon_{ij}, j = 1, \ldots, m_i, i = 1, \ldots, n, \tag{1}$$

where μ denotes a function modeling the mean progression in the total population and the subject's variation from the mean curve is split into two effects, a random (vertical) additive effect $v_i + \varepsilon_{ij}$ and a (horizontal) time shift $\boldsymbol{x}_{ij}^T\beta + z_i$. The random vertical shift consists of a time-invariant shift in cognitive scores between individuals v_i and measurement error term ε_{ij}. The horizontal shift consists of a term modeling fixed effects of covariates $\boldsymbol{x}_{ij}^T\beta$ that describe differences in disease stage and a term modeling unobserved random variation in disease stage z_i. We assume that the two random effects, z_i and v_i, are independent of each other, and that $z_i \sim \mathcal{N}(0, \tau^2)$, $v_i \sim \mathcal{N}(0, \gamma^2)$. Moreover, the measurement noise terms are assumed to be identically distributed and independent of all other effects $\varepsilon_{ij} \sim \mathcal{N}(0, \sigma^2)$. The vector $\boldsymbol{x}_{ij}$ consist of observations from p potentially time-varying predictors for the ith patient with associated coefficients β. After fitting the model, the timescale formed by the shifted time points $t_{ij} + \boldsymbol{x}_{ij}^T\beta + z_i$ form a global timescale for disease progression in the population. The different components of the nonlinear model are visualised in Figure 1 with the three columns illustrating how the model takes data defined on a timescale measured since baseline (left) and models the individual trajectories defined using a latent disease time that forms the disease continuum (right). The middle column shows the fixed-effect staging along the disease continuum modeled by baseline disease severity group.

There exist many approaches for modeling the mean progression curve μ and each will have its strengths and weaknesses. For example, modeling μ using basis function approaches, such as splines, would allow approximation of very general nonlinear progression patterns, but could also lead to overfitting. In this work, we focus on parametric mean

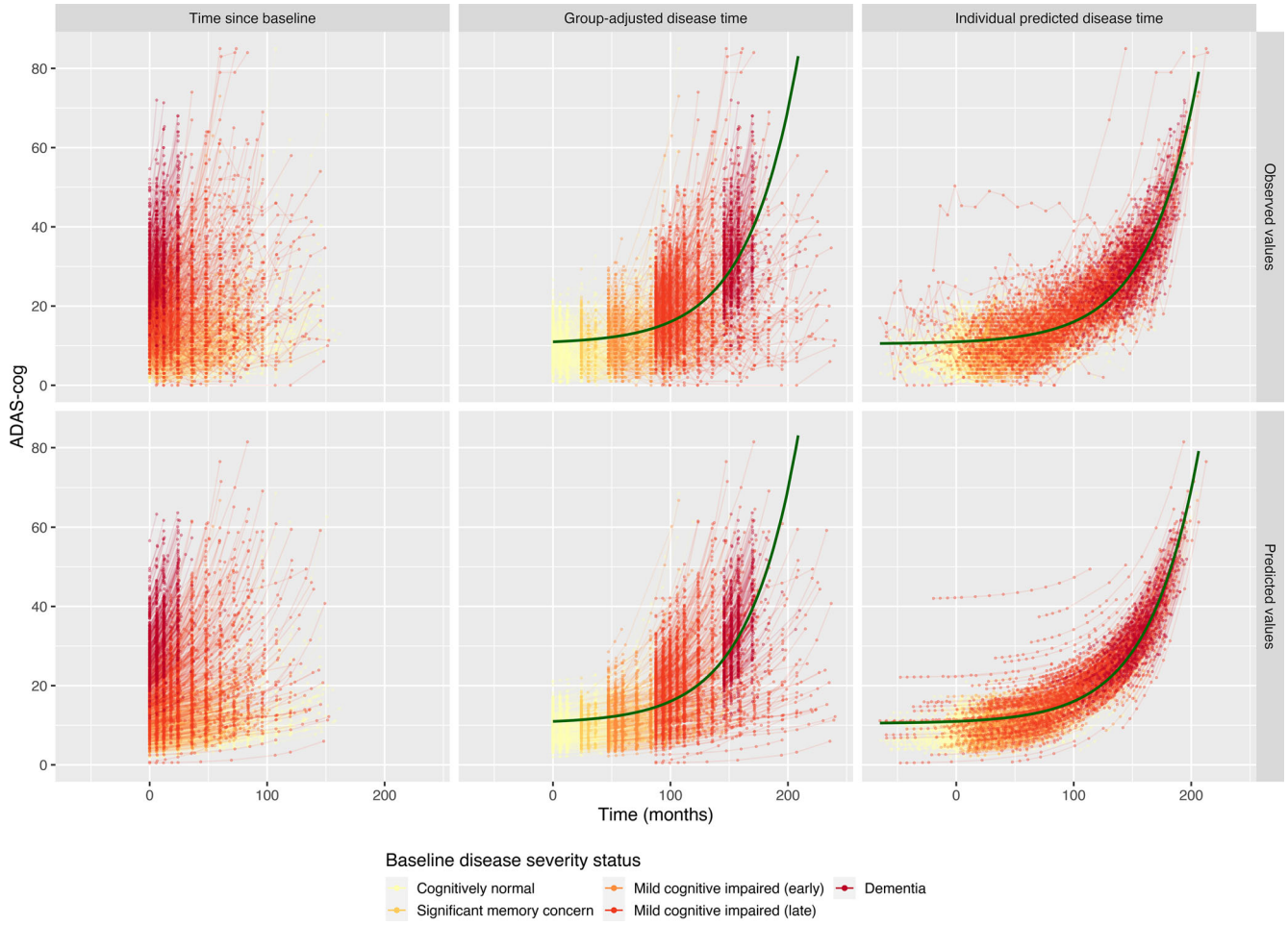

FIGURE 1 Illustration of the different components of the univariate disease progression model fitted on ADAS-cog measurements. The top row shows the observed trajectories while the bottom row shows the model's predicted subject trajectories $t \mapsto \mu(t) + v_i$ across different timescales t. Left column: Data plotted against time since study baseline. Middle column: Data plotted against group-adjusted disease time $t_{ij} + \boldsymbol{x}_{ij}^T \beta$. Right column: Data plotted against predicted disease time $t_{ij} + \boldsymbol{x}_{ij}^T \beta + z_i$. In the two latter columns, time 0 of the latent disease timescale corresponds to the average latent time of the cognitively normal group at baseline [Color figure can be viewed at wileyonlinelibrary.com]

curves because they allow more direct control of the shape of the mean function (eg, monotonicity constraints) and easier interpretation of parameters. We focus on exponential models for the mean curve of the form

$$\mu(t) = l \cdot \exp\left(\frac{t}{\exp(g)}\right) + v, \tag{2}$$

where the parameter g describes the scaling of time, v describes the left asymptote, which for cognitive outcomes can be interpreted as the mean score of cognitively normal individuals, and the parameter l describes mean deviation from v at time $t = 0$. The choice of an exponential mean curve is based on the monotonic nature of the cognitive decline in AD where the rate of deterioration is typically increasing throughout the disease. In addition, this parametric representation with only three free parameters is considerably simpler compared with other choices such as the generalized logistic function.

The model described above has a single response variable y_{ij} at every visit. However, much information is typically collected at a single visit, and since the different outcomes may measure different aspects of the disease, a multivariate model may be superior to a univariate model. For example, if one cognitive measure is sensitive early in disease while another is sensitive in the later stages of the disease, individual models for these measures may not agree on the predicted disease stages of patients and it will not be possible to compare the mean progression between the cognitive measures. This

problem can be overcome by a multivariate model for a combination of measures that define a common latent timescale across outcomes. This would enable the model to take advantage of different levels of sensitivity of the measures along the disease timeline and make a direct comparison of measures possible.

We propose the MCDP model as a multivariate extension of the model given in Equation (1). Let $k = 1, \ldots, K$ be the index for the K outcomes at a given visit. The response y_{ijk} is modeled as

$$y_{ijk} = \mu_k(t_{ij} + \boldsymbol{x}_{ij}^T\beta + z_i) + v_{ik} + \varepsilon_{ijk}, \tag{3}$$

where $z_i \sim \mathcal{N}(0, \tau^2)$, $\boldsymbol{v}_i = (v_{i1} \cdots v_{iK}) \sim \mathcal{N}(0, \Gamma)$ with an unknown covariance matrix Γ modeling the vertical correlation across time-invariant deviations in cognitive scores, and $\boldsymbol{\varepsilon}_{ij} \sim \mathcal{N}(0, \text{diag}(\boldsymbol{\sigma^2}))$ where $\boldsymbol{\sigma^2} = (\sigma_1^2 \ \ldots \ \sigma_K^2)$. Independence is assumed between z_i and $\boldsymbol{v}_i$. Note that all terms besides the latent time scale, $t_{ij} + \boldsymbol{x}_{ij}^T\beta + z_i$, depend on k.

The MCDP model and its univariate counterparts include fixed effects of the baseline disease severity status (Significant Memory Concern, Early Mild Cognitively Impaired, Late Mild Cognitively Impaired, and Dementia) on the time-shift. The corresponding parameter estimates describe the differences in disease stage of the baseline severity status groups over the continuous latent time scale. To avoid overparameterization, the time differences are modeled relative to the Cognitively Normal group at baseline. Resultingly, time 0 of the latent disease timescale corresponds to the average state of the cognitively normal participants at baseline. We have made both the univariate and multivariate models presented in Section 3 available in the *progmod* R package[20] which builds on the maximum likelihood estimation procedures in the *nlme*-package.[21] Example data and code for fitting both univariate and multivariate disease progression models are available in the package documentation. The *progmod* package also allows fitting of models with other choices of mean curves than the exponential functions considered here. All analyses were done using R 4.0.2.

2.3 | Validation

Since multiple cognitive scales are considered simultaneously, the MCDP model has access to more information to estimate the disease timeline and model the variability in data compared with corresponding univariate models. To test whether this additional information resulted in a better fit of data, the predictive performance of the models was compared. The performance of the MCDP model was compared against its univariate counterparts in terms of predicting postbaseline trajectories of the three cognitive scales using only the baseline assessment. The performance was assessed in both the training set and the held-out validation set. Predictions were only done for patients with nonmissing baseline assessments. The MCDP predictions were the maximum-a-posteriori prediction of the trained model based on each individual's baseline scores and baseline disease severity status. First, the subject-specific random vertical and horizontal shifts were predicted by optimizing the conditional posterior distribution of shifts given the baseline measures, and these were then used to predict future trajectories. Since the model specification makes it possible to predict values that exceed the limits of the cognitive scales, the predictions were censored to the boundary values of the cognitive scores if this occurred. The prediction quality was analyzed via the mean squared error (MSE) and median average deviation for both training and test data.

2.4 | Comparison to other multivariate models

The MCDP model was compared with GRACE[18] and LTJMM[22] on the three cognitive outcome measures described above as well as on simulated data. The modeling choices for GRACE and LTJMM and the simulation study are described in detail below.

While being theoretically feasible, the software packages for GRACE and LTJMM did not allow prediction for unseen individuals. Therefore, the baseline observations from the test set were included in the training set for all models. These additional baseline observations were not expected to affect the model fits in any material way since the individual random effects and their associated variance parameters can only be separated by means of longitudinal trajectories, and thus the contribution of single observation points to the likelihood or posterior function would be very limited. Furthermore, GRACE did not support participants with completely missing data on one outcome. Therefore, individuals in the test set were further required to have complete data on all three cognitive measures at baseline.

2.4.1 | Growth models by alternating conditional expectation

All outcomes were independently transformed to quantiles using a weighted percentile transformation adjusting for baseline disease severity status.[18] A nonlinear model of the form

$$\widetilde{y}_{ijk} = g_k(t_{ij} + z_i) + \alpha_{1ik} t_{ij} + \alpha_{0ik} + \varepsilon_{ijk}$$

was fitted to the transformed data $\widetilde{y}_{ijk}$ using the alternating conditional expectation approach, where $z_i \sim \mathcal{N}(0, \sigma_z^2)$ denotes the random subject-level time shift that is assumed to be independent of the random slope and intercept terms $(\alpha_{1ik}\alpha_{0ik}) \sim \mathcal{N}(0, \Sigma_j)$ and the measurement noise $\varepsilon_{ijk} \sim \mathcal{N}(0, \sigma_k^2)$. The mean progression curves g_k were modeled using monotone spline functions with 9 degrees of freedom (standard setting).

2.4.2 | Latent-time joint mixed-effects models

All outcomes were independently transformed to quantiles using a weighted quantile transformation. Subsequently, the quantiles were transformed using the inverse Gaussian quantile function.[19] These transformed outcomes were modeled as Gaussian with identity link. For comparability to the MCDP model, the model included adjustment for baseline disease severity status. The transformed outcomes $\widetilde{y}_{ijk}$ were modeled as

$$\widetilde{y}_{ijk} = \boldsymbol{x}_i^T \beta_k + \gamma_k(t_{ij} + z_i) + \alpha_{1ik} t_{ij} + \alpha_{0ik} + \varepsilon_{ijk}$$

where $z_i \sim \mathcal{N}(0, \sigma_z^2)$ denotes the random subject-level time shift, that is assumed to be independent of the random slope and intercept terms $(\alpha_{1i1}, \alpha_{0i1}, \alpha_{1i2}, \alpha_{0i2}, \alpha_{1i3}, \alpha_{0i3}) \sim \mathcal{N}(0, \Sigma)$ and the measurement noise $\varepsilon_{ijk} \sim \mathcal{N}(0, \sigma_k^2)$. The term $\boldsymbol{x}_i^T \beta_k$ models the effect of the baseline disease severity status of subject i on the transformed outcome and $\gamma_k \in \mathbb{R}$ the population slope. For all model parameters, the standard choice of weakly informative priors was used and time was modeled as years since baseline.[19]

To compare the staging of patients with the MCDP model, a subject-specific time shift that was independent of outcome and visit was defined by reparameterization of the model. The reparameterization defined the subjective time shift s_i as the average subject-level contribution of the fixed effects across outcomes plus the random subject-level time shift

$$s_i = z_i + \frac{1}{K} \sum_{k=1}^{3} \gamma_k^{-1} x_i^T \beta_k.$$

2.4.3 | Simulation study

To evaluate the models' abilities to recover subject-level shifts and thus accurately stage patients along the disease continuum, data were simulated under the maximum likelihood estimates of the MCDP models for the ADNI data. Each simulation included 500 patients from the mild cognitively impaired (late) group that had visits every 6 months for 4 years. Since only one baseline group was included, the effects related to baseline group was removed from the models.

The univariate disease progression models, the MCDP model, GRACE and LTJMM were compared in terms of the root-mean-square errors, median absolute deviations, and Spearman correlations between predicted and true subject-level shifts in 1000 simulated datasets.

2.5 | Item analysis

The estimated disease timeline from the MCDP model (3) was used to investigate differences in sensitivity of the individual items of ADAS-cog, MMSE, and CDR-SB at different disease stages. This was done to identify the most sensitive items in different stages of the disease. Such knowledge of item sensitivity can potentially be used to develop new strategies for evaluating patient cognition in clinical trials. In the analyses, items were categorized into the domains Memory, Language,

Attention, Visuo-Spatial, Motor, Executive Functioning, and Social Cognition. Many items measure ability in multiple areas simultaneously, for example, Memory and Language, so the classification was done by a clinical outcome assessment expert who gave the most weight to the primary measure of the item. Notice also that the chosen subgroups are a coarse classification of items which means that for example, the Memory group consists of items in all subcategories related to memory, for example, immediate and delayed memory tasks.

The analysis of items was based on ordinal models.[23] We used a proportional odds model assuming equality of odds ratio across item levels over a time unit. Such an assumption results in simpler interpretations of the model but it may not always be able to represent the dynamics of the observed data. Alternatively, item level specific odds ratios could be modeled, for example, by assuming monotonic increase of odds ratios for increasing item levels.[24] Due to the simpler interpretation and visualization of the model, the proportional odds model was selected for the present analysis. For each level j of item X_l, we fitted two ordinal models that estimated the probability of scoring j or lower at the continuous disease stage t,

$$\text{logit}(P(X_l \leq j)) = \theta_{lj} - \alpha_l t$$

and

$$\text{logit}(P(X_l \leq j)) = \theta_{lj} - \alpha_l t - \beta_l t^3.$$

The latter model allows more flexible modeling of the developing sensitivity of items along the disease timeline. The cubic term was included because of its monotonicity. For each item, the model with the lowest Bayesian information criterion was chosen.

For two-level items, the ordinal model reduces to logistic regression. Based on the estimated parameters; the thresholds $\boldsymbol{\theta}_l = \{\theta_{lj}\}_{j=1}^{J}$, and slopes α_l and β_l, a measure of sensitivity of each item at disease time t can be defined by computing the probability of X_l being less than or equal to j. Similar types of analyses for sensitivity of items of cognitive tests have been considered previously.[8,9] Commonly the sensitivities of items are analyzed over the range of patients' cognitive abilities, reflecting a population distribution instead of an underlying measure of progression. The difference to the presented analysis is that the sensitivity is measured over a natural timescale of disease progression obtained from the MCDP model. This allows evaluation of the sensitivity of items along the disease timeline. The ordinal models were fitted separately to each item, but the joint disease progression timescale was the same across models.

To visually compare the sensitivity of each item, we computed the intervals of time-points for which $P(X_l > j)$ was within the interval [0.1,0.5], that is, the probability of being at a level above j is between 10% and 50%. Sensitive items will have a narrow time interval for which $P(X_l > j)$ is within [0.1,0.5], as patients are more prone to rapidly jump to higher levels of the specific item.

The ordinal models were fitted using the R function *clm* from the *ordinal* package.[25]

3 | RESULTS

3.1 | Validation

To investigate the value of multivariate modeling, we compared the MCDP model of the three cognitive scales to corresponding univariate models. Figure 2 shows the predicted time shifts for each patient (random plus fixed effects) in the training set from the three univariate models of the cognitive scales plotted against each other. While the predicted disease times are strongly correlated (all Spearman correlations > 0.7) there are noticeable deviations, some of which amount to several years difference in predicted disease stage. In particular, there are differences for the cognitively normal and significant memory concern groups between ADAS-cog and the two other endpoints. The reason for this is that CDR-SB and MMSE are at around their ceilings (0 and 30, respectively) for most of these individuals throughout the follow-up period, and thus there is only very limited information available for staging individuals for these measures in the nonimpaired stages. A multivariate model could potentially utilize the different sensitivities in scores across scales to improve the predicted disease timescale.

The MCDP model presented in Section 2.2 was fitted simultaneously to assessments of ADAS-cog, MMSE, and CDR-SB in the training data. The estimated mean curve projected on to each dimension of the cognitive measures are shown in Figure 3.

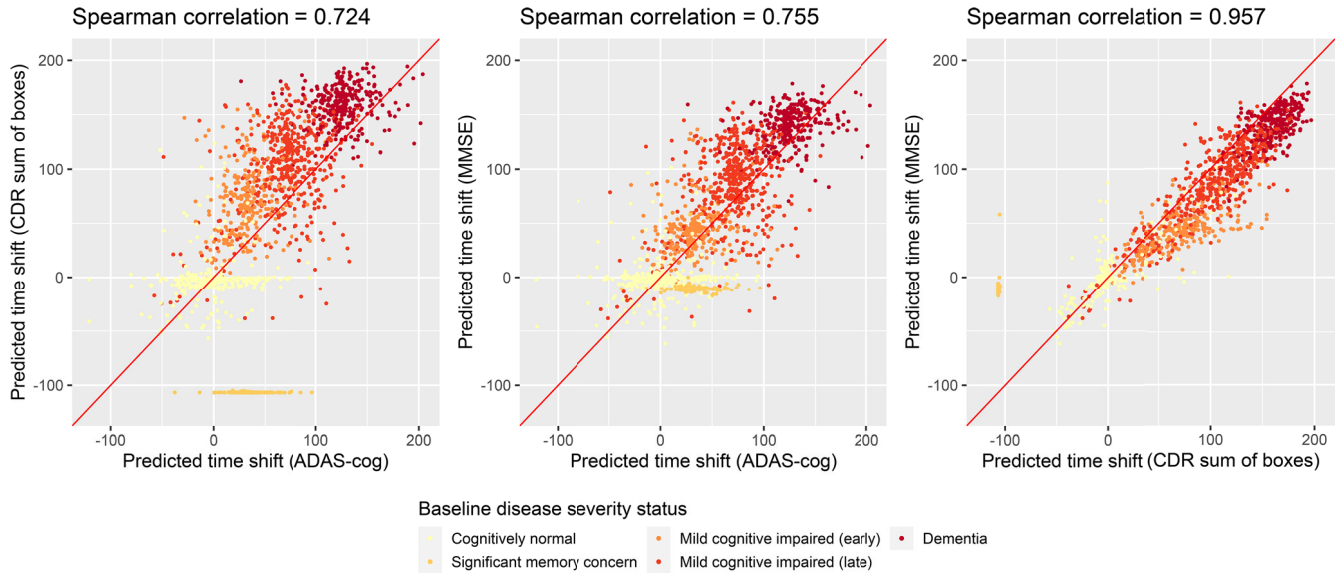

FIGURE 2 Comparison of the predicted individual time shifts from each of the three univariate models [Color figure can be viewed at wileyonlinelibrary.com]

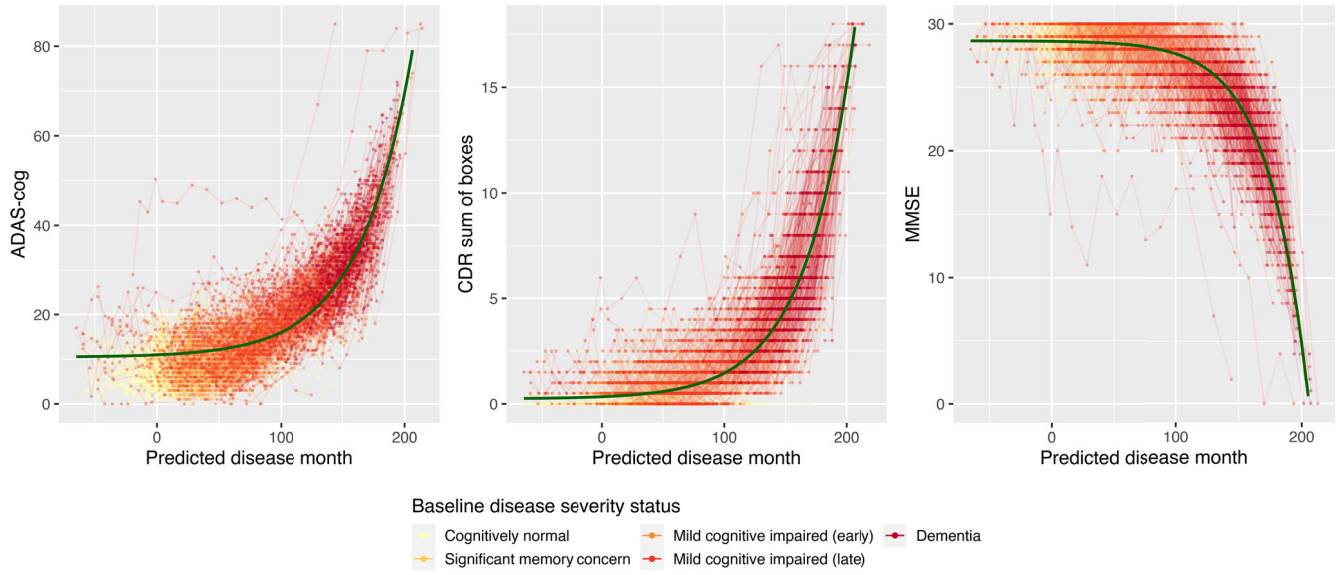

FIGURE 3 Mean curves for the three cognitive scales simultaneously estimated using the multivariate continuous-time disease progression (MCDP) model. The predicted disease month time scale is the time scale predicted by the model that is adjusted for group-level (fixed) and subject-level (random) differences in disease stage [Color figure can be viewed at wileyonlinelibrary.com]

The estimated correlations between the outcome-specific vertical shifts v_{ik} are given in Table 1. ADAS-cog and MMSE had a strong correlation of −0.7 while correlations to CDR-SB were both around 0.4 in magnitude. The SD τ for the random time shift z_i was estimated to 41.1 months, suggesting considerable disease-stage variation within the baseline disease severity groups. The Spearman correlations between the predicted disease time for the MCDP model against the predictions of each of the univariate models were all above 0.8.

The MCDP model fit was compared with each of the three univariate models described in Section 2.3. Patient-level predictions were made based on the baseline disease severity status and cognitive measures at baseline. The MSEs and MADs for the univariate models and MCDP model on each cognitive scale are given in Table 2. In the test set, the MCDP model had significantly lower MSE of postbaseline predictions of patient trajectories on ADAS-cog and MMSE (both

TABLE 1 Estimated correlations between the outcome specific vertical shift effects for the MCDP model

	ADAS-cog	CDR-SB	MMSE
ADAS-cog	1		
CDR-SB	0.394	1	
MMSE	−0.733	−0.473	1

Abbreviations: MCDP, multivariate continuous-time disease progression; MMSE, Mini-Mental State Examination.

TABLE 2 Mean squared errors and median absolute deviations of postbaseline predictions of univariate disease progression models and the MCDP model

		Mean squared error			Median absolute deviation		
		ADAS-cog	CDR-SB	MMSE	ADAS-cog	CDR-SB	MMSE
TRAIN	Univariate	142.9	7.82	23.2	4.03	**0.63**	1.40
	MCDP	**106.9**	**6.96**	**16.9**	**3.87**	0.69	**1.32**
TEST	Univariate	129.2	8.20	20.2	3.95	**0.59**	1.41
	MCDP	**100.3**	**7.02**	**16.5**	**3.85**	0.68	**1.35**

Note: Boldface indicates best performance.
Abbreviations: GRACE, growth models by alternating conditional expectation; LTJMM, latent-time joint mixed-effects models; MCDP, multivariate continuous-time disease progression; MMSE, Mini-Mental State Examination.

$P < .001$, paired Wilcoxon tests). There was no significant difference in the postbaseline predictions on CDR-SB ($P = .41$, paired Wilcoxon test).

3.2 | Comparison to other multivariate models

LTJMM, GRACE, and the MCDP model were all fitted on the joint dataset consisting of the training set and the baseline observations from the test set. There were big differences in runtimes of the algorithms. The MCDP model required approximately 2 minutes for convergence, GRACE ran for approximately 15 minutes and LTJMM ran for approximately 12 hours per chain. Figure 4 show the fitted models on each outcome along with the estimated population mean progression curves.

Sensitivity analyses that excluded the test set baseline observations confirmed that inclusion of additional baseline observations had no relevant impact on the fits. Pearson correlations between model residuals and between predicted random time shifts for the individuals in the training set were all between 0.99 and 1.00 for the three multivariate models.

The models were compared based on their predictive accuracy of postbaseline patient trajectories on the three cognitive scales. The predictive performances of the models are given in Table 3 which shows that the MCDP produces superior prediction results to the other models across all three endpoints (all $P < .0001$, paired Wilcoxon tests).

In the simulation study, the MCDP model consistently produced more accurate patient staging than LTJMM and GRACE with an average Spearman correlation across simulations of 0.93 compared with 0.52 and 0.83. The full results are shown in Figure 5. Interestingly, the three univariate models all performed better than GRACE and LTJMM with the model using ADAS-cog performing best followed by the CDR-SB model and the MMSE model.

3.3 | Scale and item sensitivities

Figure 6 shows the three projected normalized mean curves of ADAS-cog, MMSE (reversed for comparison), and CDR-SB estimated with the MCDP model. Due to differences in range and variation of the cognitive scores, the mean curves and their derivatives were divided by the estimated SD of the measurement noise in each dimension of the MCDP model.

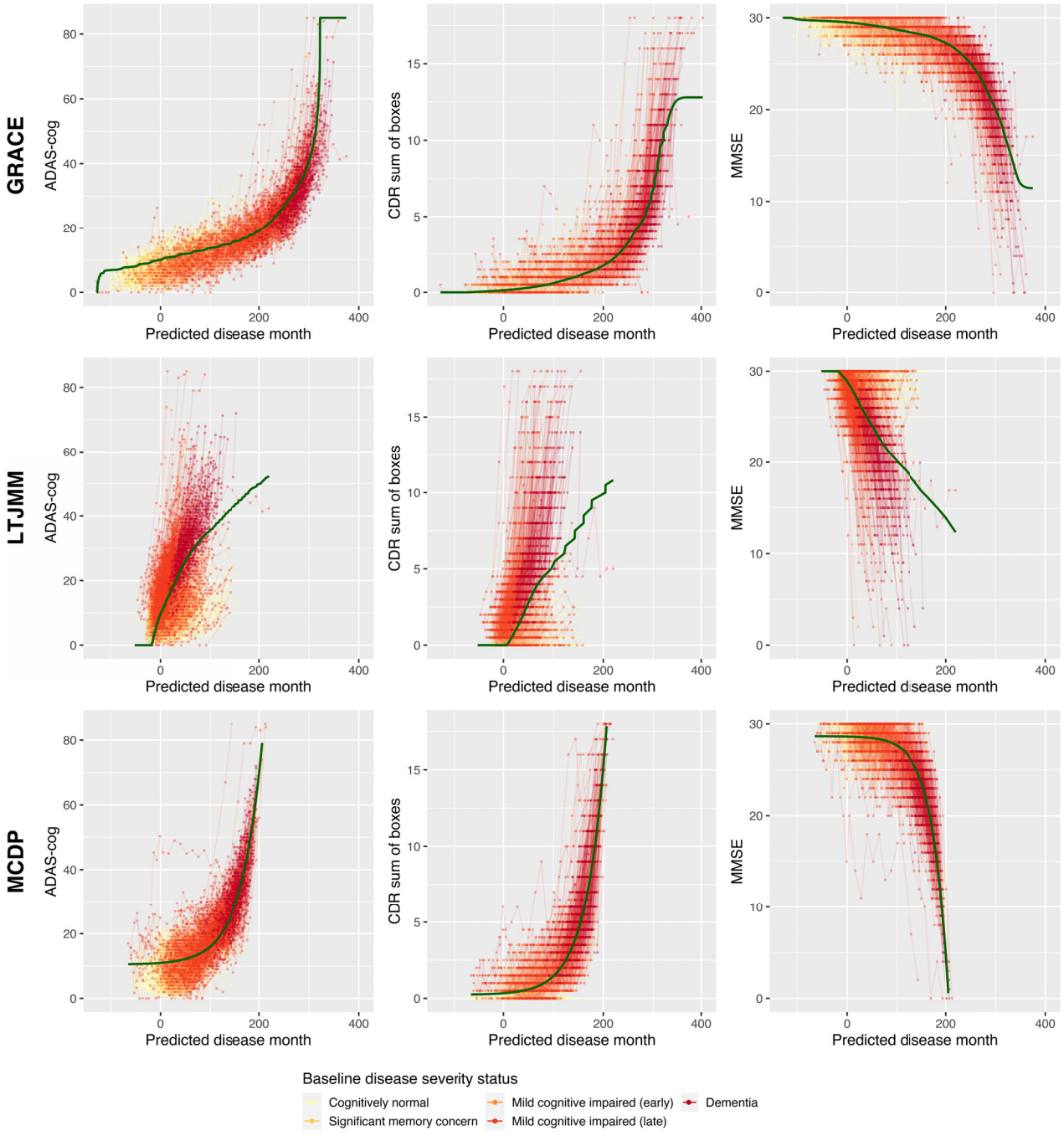

FIGURE 4 Comparison of estimated mean curves and predicted patient staging across models. For better comparison, the GRACE and LTJMM results have been shifted such that time zero corresponds to the average predicted disease month at baseline for Cognitively normal individuals. For LTJMM, the slight residual variation from the intercepts of the baseline disease severity groups is not shown. GRACE, growth models by alternating conditional expectation; LTJMM, latent-time joint mixed-effects models [Color figure can be viewed at wileyonlinelibrary.com]

In addition, the baseline level was subtracted from the mean curves to make a better comparison of the evolution of curves. From the plot of the absolute derivatives, Figure 5 (right), it is seen that ADAS-cog and CDR-SB have similar early increases on the predicted disease timescale, while MMSE has a steeper increase in the later stages of the disease. This suggests that ADAS-cog and CDR-SB are more sensitive measures for detecting changes early in the disease, while MMSE becomes the most sensitive of the three later in the disease.

Based on the predicted disease month for each subject obtained from the MCDP model, we compared the sensitivity of different items included in ADAS-cog, MMSE, and CDR. The aim was to identify the items with the steepest increase at

TABLE 3 Mean squared errors and median absolute deviations of postbaseline predictions on the test set

	Mean squared error			Median absolute deviation		
	ADAS-cog	**CDR-SB**	**MMSE**	**ADAS-cog**	**CDR-SB**	**MMSE**
GRACE	114.0	12.5	24.1	4.28	0.877	1.53
LTJMM	148.7	8.17	18.5	7.78	1.48	2.48
MCDP	**103.4**	**7.18**	**17.1**	**3.83**	**0.667**	**1.34**

Note: Boldface indicates best performance.
Abbreviations: GRACE, growth models by alternating conditional expectation; LTJMM, latent-time joint mixed-effects models; MCDP, multivariate continuous-time disease progression; MMSE, Mini-Mental State Examination.

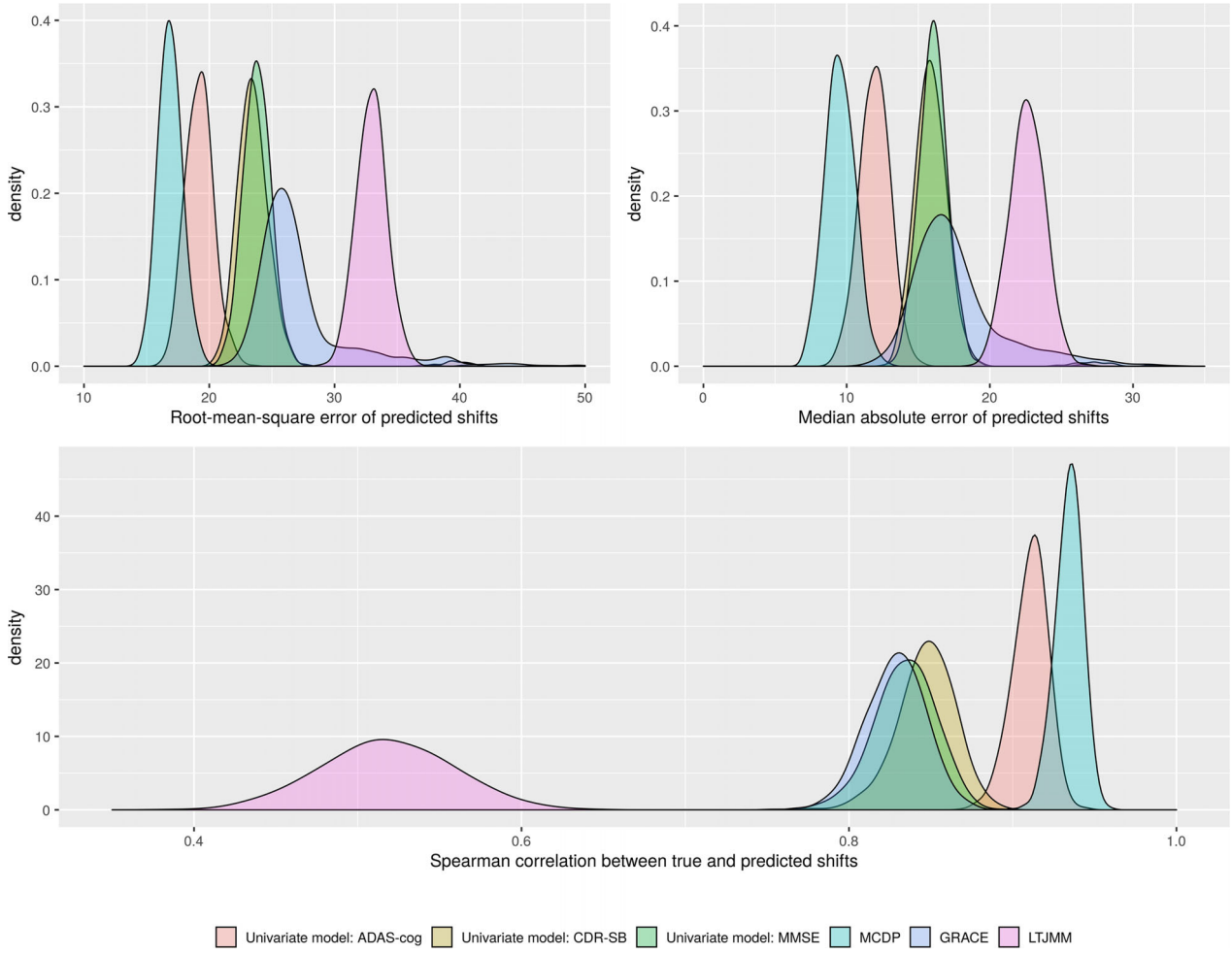

FIGURE 5 Density plots of the root-mean-square errors, median absolute errors, and Spearman correlations between the true and predicted shifts for univariate disease progression models, MCDP, LTJMM, and GRACE across 1000 simulations. GRACE, growth models by alternating conditional expectation; MCDP, multivariate continuous-time disease progression; LTJMM, latent-time joint mixed-effects models [Color figure can be viewed at wileyonlinelibrary.com]

different time intervals of the predicted disease timeline to better understand the continuous evolution of patient symptoms and aid the development of new and more sensitive cognitive assessments. Two item scores from each cognitive outcome measure are plotted against predicted disease month and shown in Figure 7. As can be seen from the figure, some items show a clearer evolution over the predicted disease timescale than others and would hence be more informative for describing the degree of cognitive impairment of a subject.

To analyse differences in sensitivity of items, we fitted ordinal models to each item of the three cognitive scales as described in Section 2.5. Figures 8 and 9 show the results of the ordinal analysis for all three cognitive scores. Items

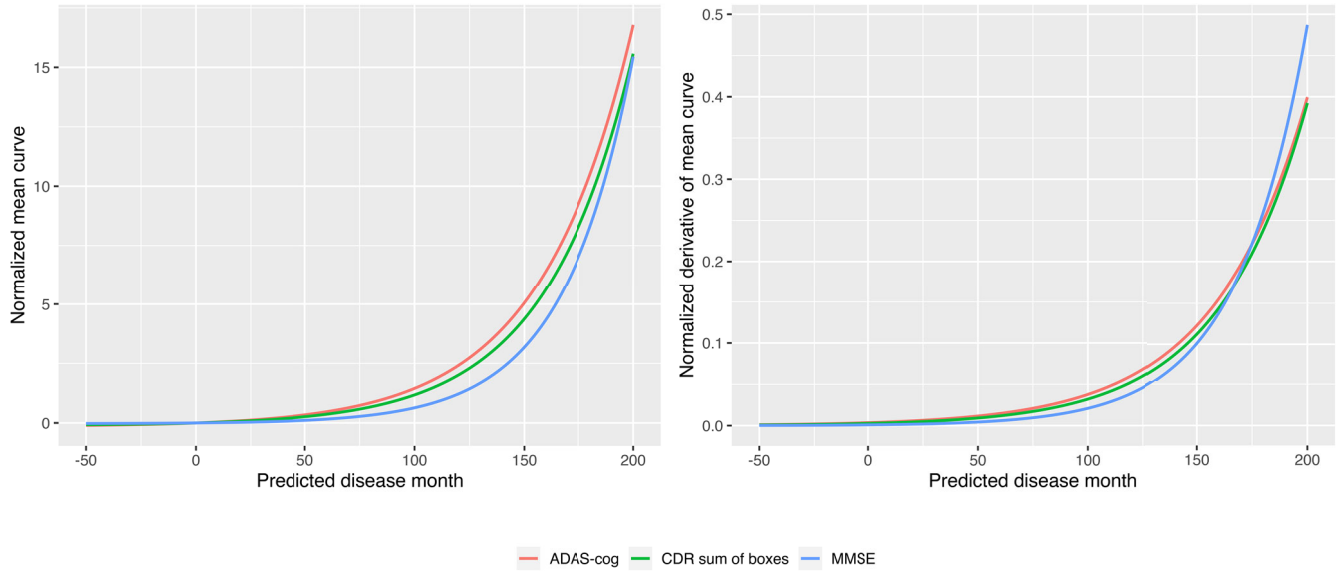

FIGURE 6 Comparison of normalized mean curves (left) and their corresponding absolute derivatives (right). The curves were normalized with the estimated SD of the measurement noise of each cognitive measure, and the mean curves were shifted vertically to have a common starting value at zero [Color figure can be viewed at wileyonlinelibrary.com]

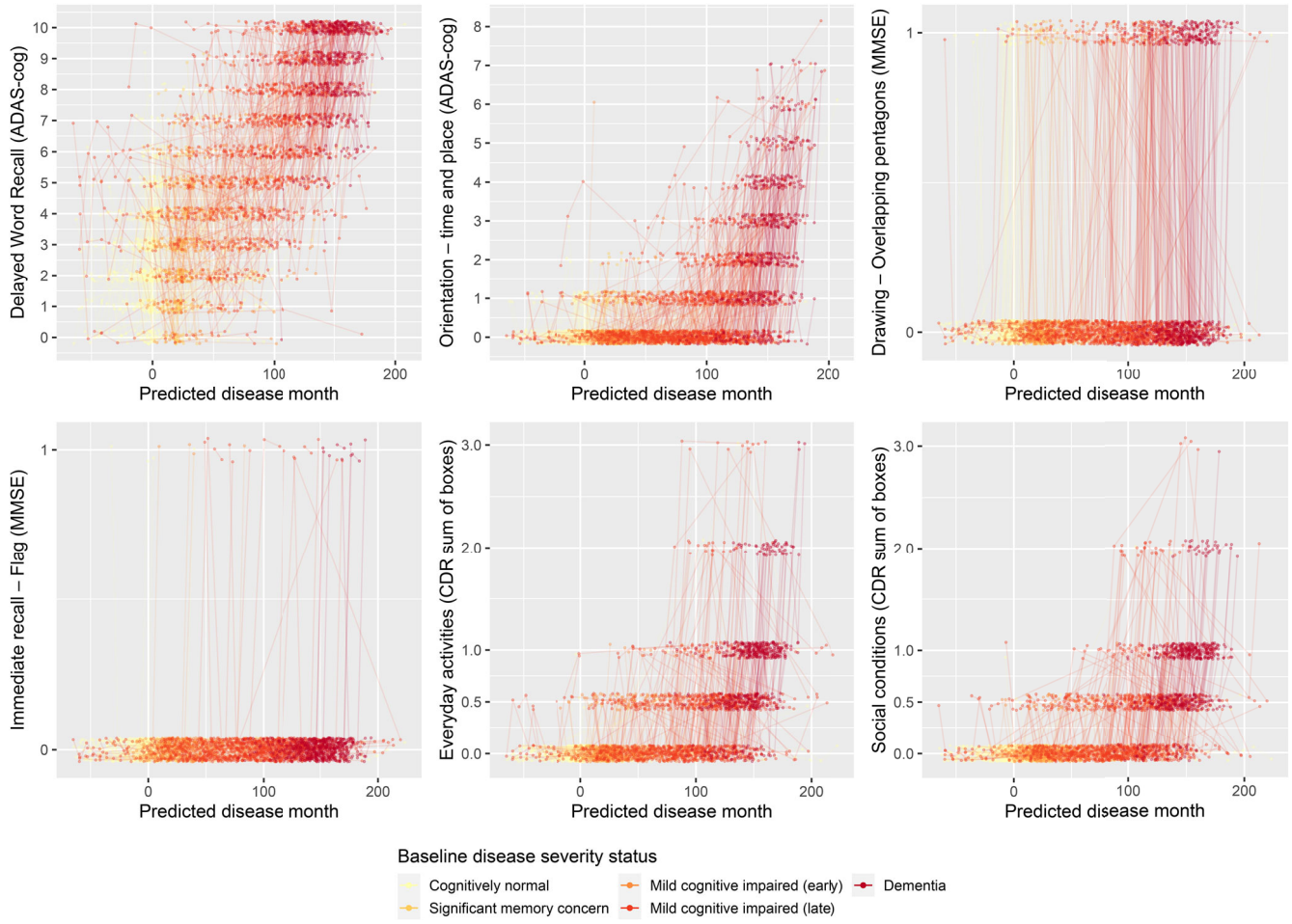

FIGURE 7 Examples of individual item score trajectories plotted against predicted disease month found with the multivariate continuous-time disease progression (MCDP) model [Color figure can be viewed at wileyonlinelibrary.com]

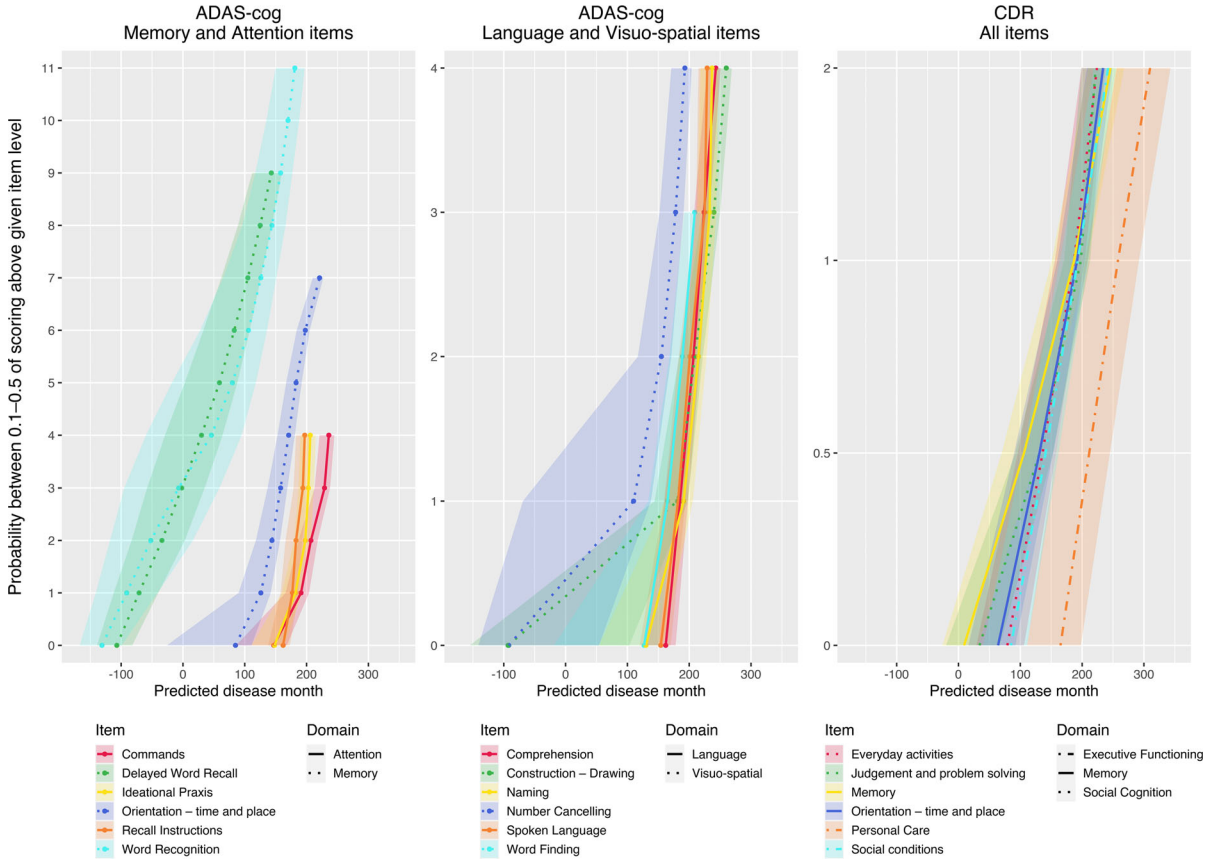

FIGURE 8 Sensitivity of ADAS-cog and CDR items across the predicted disease time. The visualized intervals for sensitivity comparison of items is based on the timepoints for which the probability of belonging to a level greater than the current one is between 10% and 50%. The lines within intervals represent the 30% threshold. Narrow time intervals suggest higher sensitivity of items while broad intervals suggest less information for predicting cognitive severity of patients [Color figure can be viewed at wileyonlinelibrary.com]

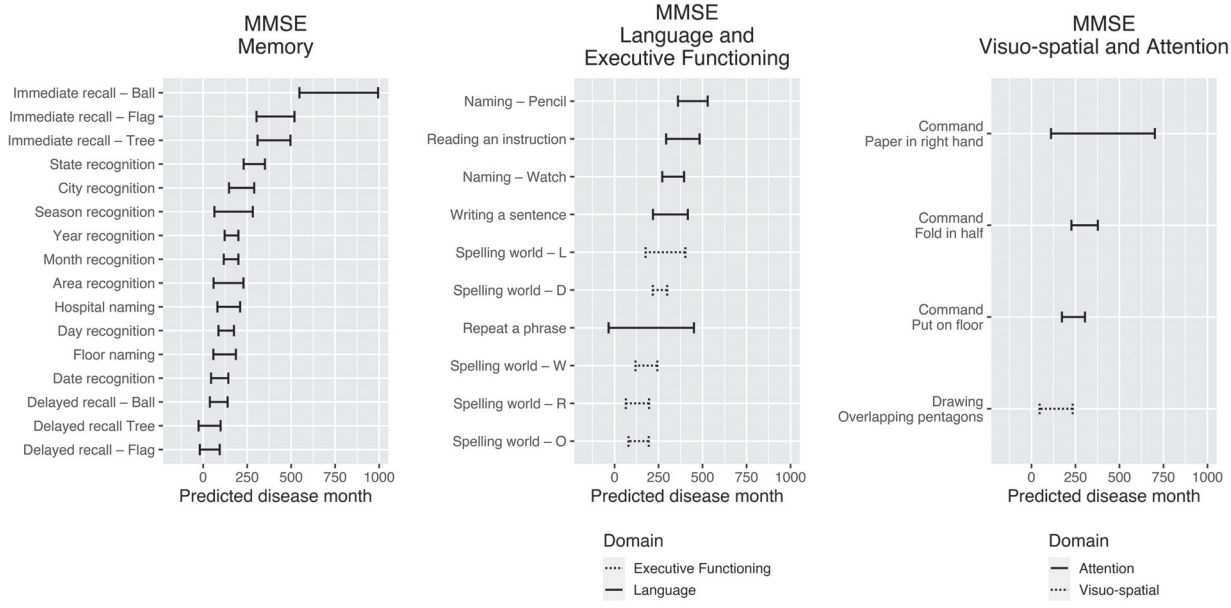

FIGURE 9 Sensitivity of Mini-Mental State Examination (MMSE) items across the predicted disease time. The visualized intervals for sensitivity comparison of items is based on the timepoints for which the probability of answering incorrect is within 10% to 50%. All items are for the MMSE score. Narrow time intervals suggest higher sensitivity of items while broad intervals suggest less information for predicting cognitive severity of patients

with narrow intervals at the beginning of the predicted timeline would be good options to include in an instrument for measuring progression in the early stages of AD, while wide-banded items that span long time intervals late in the disease would be less informative. The first item of the ADAS-cog, Word Recall, was omitted from the analysis since a subset of the scores were measured as an average of nonrecalled words over three-word lists while others only used a single word list.

We found that the items in the Memory domain were the most sensitive in the early stages of the disease. The most sensitive memory items highlighted by the analyses for both MMSE and ADAS-cog were delayed memory recall tests. These items show an early narrow sensitivity interval in the ordinal analyses. Besides the memory tests, items showing an early sensitivity include tests of orientation to time and place and tests concerning the comprehension of everyday activities and social life. In the later stage of the disease, items testing mostly language and attention become sensitive, while the floor is reached for the memory tests which thus hold little information on the progression in these late stages of the disease.

4 | DISCUSSION

Staging of AD patients is often considered a discrete problem with stages including preclinical AD, mild cognitive impairment, and mild/moderate/severe dementia. The presented MCDP model challenges the discrete representation of the progression of subjects. Through nonlinear mixed-effects modeling, we demonstrated the possibility of estimating a continuous trajectory representing the progression of AD and to predict disease stage of individual patients. The MCDP model was used on data from the ADNI study to simultaneously model the three cognitive scales ADAS-cog, MMSE, and CDR-SB. The model was shown to align patients along the disease trajectory resulting in a marked decrease in variation of cognitive measure compared with analysis of the measures using time since study inclusion. Furthermore, it was seen that patients classified in the same discrete baseline disease severity groups could have substantial differences in predicted disease stage.

Compared with univariate modeling, the MCDP model was shown to be significantly better in predicting longitudinal observations on the individual scales of unseen individuals on ADAS-cog and MMSE, while no significant difference in predictive performance was found on CDR-SB.

The MCDP model was subsequently compared with two other disease progression models: LTJMM and GRACE. Comparing the predicted staging and estimated mean progression curves for the MCDP model, LTJMM and GRACE revealed that the models produced substantial differences in the estimated duration of disease and in progression speed. GRACE found slower disease progression and longer overall disease duration, compared with the MCDP model. LTJMM was seen to model the main variation in the population by the random vertical intercept and slope, leaving almost no variation to the disease stage effect. When compared in terms of predictive performance on held-out data, the MCDP model was shown to be significantly better at predicting future cognitive decline compared with the two other multivariate models.

Next, the univariate models, the MCDP model, LTJMM, and GRACE were compared in a simulation study to evaluate their ability to recover individual patient stages along the disease trajectory. The results showed clearly superior performance of the MCDP model, but also better performance of the three univariate models compared with LTJMM and GRACE. The simulation scenario may have favored the MCDP and univariate models because they were all correctly specified for the data. Simulations under the estimates of LTJMM and GRACE were not considered feasible because both methods require preprocessing of data including nonlinear transformations (including the rank transform). Thus, there is no natural way to back-transform such simulated data to the original scales of the cognitive outcomes that the MCDP model was developed to model. LTJMM and GRACE rely on transformation of data for handling different types of outcomes, while different outcomes require explicit modeling choices of mean trajectories for MCDP (several such choices are available in the accompanying *progmod* R package). As a result, LTJMM and GRACE may be able to model many different types of data more robustly and with less explicit modeling choices than MCDP. However, the poor performance of LTJMM suggests that this approach may in general not be very robust.

Taken together, these findings suggest that the progression curves estimated with the MCDP model better represent the true evolution of cognitive scores in AD compared with the alternative models and that the model is better to predict individual patient's disease progression.

With the continuous disease staging predicted by the MCDP model, we finally analyzed individual item sensitivities along the estimated disease timeline. Items from the three cognitive scales were compared via an ordinal model

framework describing the probability of staying at an item level. Such multivariate analysis gives a unique possibility of comparing items across cognitive measures. The ordinal analysis showed that delayed memory items of all cognitive measures were sensitive early in the disease. In later stages of the disease, items mostly measuring language and attention difficulties stood out as sensitive, while the memory items had reached their floor. The continuous evolution of item scores over disease time makes it possible to not only describe which domains are affected first but also when and for how long into disease a specific item is sensitive.

The developed methodology and studies presented here have some limitations. First, the MCDP model assumes that missing data are missing at random, which implies that missingness can be fully accounted for by observed data. This assumption is often violated when working with longitudinal data from elderly individuals, where factors such as age-related disease and mortality may lead to changing populations over time.[26] In AD studies, in particular, individuals are more likely to drop out as the severity of disease increases.[27] This may, in turn, have biased estimates, however, this effect would most likely be greatest at the late stages of the disease where data is limited. Furthermore, the MCDP model assumes that all individuals fit on the same trajectory, however, some individuals may never develop AD or develop other diseases with distinct cognitive trajectories. This leads to the model shifting the cognitively normal individuals who do not experience pathological ageing further back in predicted disease time as observation time increases. While this should have limited impact on the overall results, this will weaken the predictive performance of the model when predictions are made for cognitively normal individuals based on limited data (eg, baseline assessments). This problem can likely be lessened by including sensitive biomarker outcomes in the modeling, but future work should address this properly by directly modeling normal cognitive and pathological trajectories and the transitions between the two.

ACKNOWLEDGEMENTS

This work is partly funded by the Innovation Fund Denmark (IFD) under File No. 9066-00005B. 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). 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.

DATA AVAILABILITY STATEMENT

Data used in 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 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.

ORCID

Line Kühnel https://orcid.org/0000-0003-4797-9554
Lars L. Raket https://orcid.org/0000-0001-7099-2314

REFERENCES

1. Alzheimer's Association. 2019 Alzheimer's disease facts and figures. *Alzheimers Dement*. 2019;15(3):321-387.
2. Rosen WG, Mohs RC, Davis KLA. New rating scale for Alzheimer's disease. *Am J Psychiatry*. 1984;141(11):1356-1364.
3. Zhang D, Wang Y, Zhou L, Yuan H, Shen D. Multimodal classification of Alzheimer's disease and mild cognitive impairment. *Neuroimage*. 2011;55(3):856-867.

4. Mattsson N, Palmqvist S, Stomrud E, Vogel J, Hansson O. Staging β-amyloid pathology with amyloid positron emission tomography. *JAMA Neurol*. 2019;76(11):1319–1329.
5. Insel PS, Weiner M, Mackin RS, et al. Determining clinically meaningful decline in preclinical Alzheimer disease. *Neurology*. 2019;93(4):322-333.
6. Soto ME, Andrieu S, Arbus C, et al. Rapid cognitive decline in Alzheimer's disease. consensus paper. *J Nutr Health Aging*. 2008;12(10):703-713.
7. Snyder PJ, Kahle-Wrobleski K, Brannan S, et al. Assessing cognition and function in Alzheimer's disease clinical trials: do we have the right tools? *Alzheimers Dement*. 2014;10(6):853-860.
8. 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.
9. Ashford JW, Kolm P, Colliver JA, Bekian C, Hsu L-N. Alzheimer patient evaluation and the mini-mental state: item characteristic curve analysis. *J Gerontol*. 1989;44(5):139-146.
10. Gauthier S, Vellas B, Farlow M, Burn D. Aggresive course of disease in dementia. *Alzheimers Dement*. 2006;2(3):210-217.
11. Ballard C, Atri A, Boneva N, et al. Enrichment factors for clinical trials in mild-to-moderate Alzheimer's disease. *Alzheimer's Dementia Transl Res Clin Intervent*. 2019;1(5):164-174.
12. Adak S, Illouz K, Gorman W, et al. Predicting the rate of cognitive decline in aging and early Alzheimer disease. *Neurology*. 2004;63(1):108-114.
13. Kester MI, van der Vlies AE, Blankenstein MA, et al. CSF biomarkers predict rate of cognitive decline in Alzheimer disease. *Neurology*. 2009;73(17):1353-1358.
14. Stern Y. Cognitive reserve in ageing and Alzheimer's disease. *Lancet Neurol*. 2012;11(11):1006-1012.
15. Andel R, Vigen C, Mack WJ, Clark LJ, Gatz M. The effect of education and occupational complexity on rate of cognitive decline in Alzheimer's patients. *J Int Neuropsychol Soc*. 2006;12(1):147-152.
16. Cnaan A, Laird NM, Slasor P. Using the general linear mixed model to analyse unbalanced repeated measures and longitudinal data. *Stat Med*. 1997;16(20):2349-2380.
17. Jedynak BM, Lang A, Liu B, et al. A computational neurodegenerative disease progression score: method and results with the Alzheimer's Disease Neuroimaging Initiative cohort. *Neuroimage*. 2012;63(3):1478-1486.
18. Donohue MC, Jacqmin-Gadda H, Le Goff M, et al. Estimating long-term multivariate progression from short-term data. *Alzheimers Dement*. 2014;10:400-410.
19. Li D, Iddi S, Thompson WK, Donohue MC. Bayesian latent time joint mixed effect models for multicohort longitudinal data. *Stat Methods Med Res*. 2019;28(3):835-845.
20. Raket LL. Statistical disease progression modeling in Alzheimer disease. *Front Big Data*. 2020;3:1-18.
21. Pinheiro JC, Bates DM. Mixed-effects models in S and S-PLUS. *Statistics and Computing*. New York, NY: Springer; 2000.
22. Li D, Iddi S, Thompson WK, Donohue MC. For the Alzheimers disease neuroimaging initiative, Bayesian latent time joint mixed effect models for multicohort longitudinal data. *Stat Methods Med Res*. 2019;28(3):835-845.
23. McCullagh P. Regression models for ordinal data. *J R Stat Soc B Methodol*. 1980;42(2):109-142.
24. Capuano AW, Dawson JD. The trend odds model for ordinal data. *Stat Med*. 2013;32(13):2250-2261.
25. Christensen RHB. Ordinal - regression models for ordinal data. R package version 2019.4–25; 2019.
26. Harel O, Hofer SM, Hoffman L, Pedersen NL, Johansson B. Population inference with mortality and attrition in longitudinal studies on aging: a two-stage multiple imputation method. *Exp Aging Res*. 2007;33(2):187-203.
27. William-Faltaos D, Chen Y, Wang Y, Gobburu J, Zhu H. Quantification of disease progression and dropout for Alzheimer's disease. *Int J Clin Pharmacol Ther*. 2013;51(2):120-131.

How to cite this article: Kühnel L, Berger A-K, Markussen B, Raket LL. Simultaneous modeling of Alzheimer's disease progression via multiple cognitive scales. *Statistics in Medicine*. 2021;1–16. https://doi.org/10.1002/sim.8932