J Pharmacokinet Pharmacodyn (2012) 39:479–498
DOI 10.1007/s10928-012-9263-3

ORIGINAL PAPER

Combining patient-level and summary-level data for Alzheimer's disease modeling and simulation: a beta regression meta-analysis

James A. Rogers · Daniel Polhamus · William R. Gillespie · Kaori Ito · Klaus Romero · Ruolun Qiu · Diane Stephenson · Marc R. Gastonguay · Brian Corrigan

Received: 14 February 2012 / Accepted: 3 July 2012 / Published online: 21 July 2012

Abstract Our objective was to develop a beta regression (BR) model to describe the longitudinal progression of the 11 item Alzheimer's disease (AD) assessment scale cognitive subscale (ADAS-cog) in AD patients in both natural history and randomized clinical trial settings, utilizing both individual patient and summary level literature data. Patient data from the coalition against major diseases database (3,223 patients), the Alzheimer's disease neruroimaging initiative study database (186 patients), and summary data from 73 literature references (representing 17,235 patients) were fit to a BR drug-disease-trial model. Treatment effects for currently available acetyl cholinesterase inhibitors, longitudinal changes in disease severity, dropout rate, placebo effect, and factors influencing these parameters were estimated in the model. Based on predictive checks and external validation, an adequate BR meta-analysis model for ADAS-cog using both summary-level and patient-level data was developed. Baseline ADAS-cog was estimated from baseline MMSE score. Disease progression was dependent on time, ApoE4 status, age, and gender. Study drop out was a function of time, baseline age, and baseline MMSE. The use of the BR constrained simulations to the 0–70 range of the ADAS-cog, even when residuals were incorporated. The model allows for simultaneous fitting of summary and patient level data, allowing for integration of all information available. A further advantage of the BR model is that i t constrains values to the range of the original instrument for simulation purposes, in contrast to methodologies that provide appropriate constraints only for conditional expectations.

Keywords Alzheimer's disease · Disease progression · Beta regression · Meta-analysis · Logistic · Longitudinal · Bayesian · CAMD · ADNI

This study is conducted for the Coalition Against Major Diseases and the Alzheimer's Disease Neuroimaging Initiative.

Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.ucla.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.ucla.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf. Data used in the preparation of this article were obtained from the Coalition Against Major Diseases database (CAMD).

J. A. Rogers · D. Polhamus · W. R. Gillespie · M. R. Gastonguay
Metrum Research Group, Tariffville, CT 06081, USA

K. Ito · R. Qiu · B. Corrigan (✉)
Pfizer Global Research Group, 445 Eastern Point Road, Groton, CT 06340, USA
e-mail: brian.corrigan@pfizer.com

K. Romero · D. Stephenson
Critical Path Institute, 1730 East River Road, Tucson, AZ 85718, USA

Introduction

The Alzheimer's disease assessment scale (ADAS) was designed to measure the severity of the most important symptoms of Alzheimer's disease (AD) [17]. Its cognitive subscale, ADAS-cog is the *de facto* standard primary outcome neuropsychological measure for AD trials [22]. It consists of 11 tasks measuring the disturbances of memory, language, praxis, attention and other cognitive abilities that are often referred to as the core symptoms of AD. It has been extensively validated in English as well as numerous

other languages. ADAS-cog scores range from 0 to 70, with higher scores indicating greater cognitive impairment. Our present objective is to develop a regression model that describes disease progression as measured by ADAS-cog, in AD patients. More specifically, we aim to provide a regression model that permits a comprehensive synthesis of available evidence, while also permitting the simulation of realistic patient-level data.

The bounded nature of the ADAS-cog implies, *a fortiori*, nonlinear progression. In studies of normal elderly and early AD populations that utilize the ADAS-cog, it is not unusual to observe some patients with scores consistently near the floor of the instrument (0), while in studies of longer duration, moderate AD patients may progress to scores near the ceiling of the instrument (70) over the course of a study. Consequently, a rigorous approach to the modeling of bounded scores is called for. These challenges have been partially addressed by Ashford and Schmitt [2] and subsequently by Samtani et al [23]. Specifically, their models provide appropriate constraints for conditional expectations ranging from 0 to 70 and support valid inferences regarding population means. Despite these substantial modeling advancements, Ashford and Schmitt did not provide stochastic structures that could be used to simulate new data, and the stochastic elements utilized by Samtani et al. result in predictive distributions that are not fully constrained to the 0–70 range.

Kieschnick and McCullough [11] provide an excellent review of seven possible strategies for analyzing constrained response data. Included among the reviewed methods were: approaches involving transformation of the dependent variable, approaches using nonlinear regression with normal residuals, approaches based on the censored normal distribution, and approaches based on the beta regression. The term "beta regression" has been used to refer to regression (typically logistic regression) under the assumption that residuals follow a beta distribution. This methodology has been advocated in the statistics literature as an effective means of modeling responses with constrained scales.

A beta regression model is proposed to describe the longitudinal progression of the ADAS-cog, in AD patients in both natural history and randomized clinical trial settings. The advantage of the beta regression methodology in this context is that the entire predictive distribution is limited to the desired constraints, in contrast to methodologies that provide appropriate constraints only for conditional expectations. This feature is particularly advantageous in the context of clinical trial simulation, since realistic data are generated even when residuals are incorporated.

Also described in this paper is a modified version of the usual beta regression model that is appropriate for meta-analyses, including those based on a combination of summary-level and patient-level data. To our knowledge, adaptations of beta regression to meta-analysis of this type have not previously been published. The resultant model structure described is a complete drug-disease trial model for ADAS-cog that incorporates all available information, from clinical trial and literature sources, and that is suitable for clinical trial simulation purposes. The methodology is broadly applicable to other endpoints.

Methods

Data

Data from three sources was utilized for model development and testing: (1) The AD neuroimaging initiative trial (ADNI), (2) The coalition against major diseases (CAMD) database, and (3) the literature. ADNI data provided a rich source for the natural longitudinal history of disease progression in patients with mild and moderate AD, and one of the most complete sources of imaging and biomarker data collected to date in any AD observational trial. The CAMD database provided a rich standardized source for individual level control arm data in mild and moderate AD patients (both placebo and stable background therapy) from randomized controlled clinical trials (refer to Table 2). The literature (summary level data) provided data for the model to estimate symptomatic treatment effects for AChE inhibitors, long term disease progression in controlled mild and moderate AD trials, and inter study variability. Each source is described below in more detail.

ADNI

Data used in the preparation of this article were obtained from the Alzheimer's disease neuroimaging initiative (ADNI) database (adni.loni.ucla.edu). The ADNI was launched in 2003 by the National Institute on Aging (NIA), the National Institute of Biomedical Imaging and Bioengineering (NIBIB), the Food and Drug Administration (FDA), private pharmaceutical companies and non-profit organizations, as a $60 million, 5-year public-private partnership. The primary goal of ADNI has been to test whether serial magnetic resonance imaging (MRI), positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment (MCI) and early AD. Determination of sensitive and specific markers of very early AD progression is intended to aid researchers and clinicians to develop new treatments and monitor their effectiveness, as well as lessen the time and cost of clinical trials.

The Principal Investigator of this initiative is Michael W. Weiner, MD, VA Medical Center and University of California San Francisco. ADNI is the result of efforts of many co-investigators from a broad range of academic institutions and private corporations, and subjects have been recruited from over 50 sites across the U.S. and Canada. The initial goal of ADNI was to recruit 800 adults, ages 55–90, to participate in the research, ~200 cognitively normal older individuals to be followed for 3 years, 400 people with MCI to be followed for 3 years and 200 people with early AD to be followed for 2 years. For up-to-date information, see http://www.adni-info.org.

For this analysis, only data from the AD cohort was used. AD patients were assessed at 0, 6, 12, and 24 months. Detailed protocol information can be found at http://www.adni-info.org.

CAMD

Led by Critical Path Institute (C-Path), CAMD is a consortium of industry, FDA, EMA and patient advocacy groups in Alzheimer's and Parkinson's diseases. The mission of CAMD is to create drug development tools (biomarkers, disease models, etc) and advance them to regulatory authorities for review and approval, through sharing of non-competitive information across stakeholders [21]. As part of its deliverables, CAMD developed a standardized database of patient-level control arm data from legacy clinical trials in AD, provided by CAMD member companies. Existing standards set by the clinical data interchange standards consortium (CDISC) were used, and new ones were created wherever current standards did not yet exist (http://www.cdisc.org/content2927) [20]. The current database has patient level information from 6,000 individual subjects, spanning the continuum of MCI and mild and moderate AD trials. Data from 3,179 mild and moderate AD patients from the CAMD database was used for this analysis. Detailed information can be found at http://www.c-path.org/CAMD.cfm.

Literature data

The following criteria were used for determining which data were included or excluded: (1) if the same results were reported in different literature sources (e.g. one was the original paper and the other was the same data from a review article), only the primary source was used; (2) if more than one summary value was reported with different statistical analyses or method of imputation for missing data at the same time point, such as observed cases (OC) and last observation carried forward (LOCF), only one value was chosen. OC was preferred over LOCF if available. In some articles, however, LOCF was used for all evaluation time points (longitudinal) and both LOCF and OC were reported at the end of the study to compare the values. In this situation, LOCF was selected for all evaluation time points within the article. Also, summary values based on completers were excluded; (3) exploratory studies were excluded if they were an open study with equal to or less than 20 patients per treatment arm; and (4) a study was excluded if the patient population was considered inappropriate (e.g. patients who dropped from a previous study). Summary information for these literature data have been previously tabulated by Ito et al. [9].

Model development

Given the success of previously published models in characterizing many aspects of the progression of ADAS-cog, key learnings from existing models were adapted in a manner that would support a comprehensive meta-analysis and that would enable realistic clinical trial simulation. A large number of features of previously published models were taken as starting points and were revisited only to the extent required to obtain satisfactory model diagnostics. These "accepted structural features" included:

1. The use of a generalized logistic function to describe the natural progression of the disease on a constrained scale [23].
2. The use of a Bateman-type function to describe the incremental placebo [8, 10].
3. The use of Emax and log-linear functions to describe the incremental effects of approved AChE inhibitors as a function of dose and time [7, 9].
4. The placement of candidate covariate effects in the model. Specifically, the use of baseline severity as a covariate on the model intercept, and the use of baseline severity, APOE genotype, and baseline age as covariates on rate of progression [10, 23].
5. The use of baseline age and baseline severity as covariates on the hazard of drop-out [26].

In addition, a number of important innovations were also implemented:

1. A Bayesian implementation has been developed, allowing for a probabilistically correct synthesis of literature meta-data with patient-level data. This allows for a particularly comprehensive analysis, leveraging all available data.
2. The generalized logistic function for expected disease progression is used in conjunction with Beta-distributed residuals (i.e. "beta regression"), resulting in a predictive distribution that falls entirely within the allowable range of ADAS-cog scores (0–70) during simulation.

3. The covariance structure is extended to include inter-study variation in intercepts and rates of progression (beyond the variation already reflected by measured study-level covariates).
4. The covariance structure is extended to include inter-study heterogeneity in variance components. This allows the model to account for the likely scenario that studies differ in the quality of the methods and investigators (potentially resulting in residual distributions with different variances in different studies) and differ as well in the diversity of the enrolled patient populations (potentially resulting in different inter-subject variances in different studies).

Model specification

Parameterizing the beta distribution for beta regression

The density of a beta distribution is typically parameterized as:

$$f(x) = x^{\alpha-1}(1-x)^{\beta-1}/B(\alpha,\beta), \quad 0 \le x \le 1$$

where $\alpha > 0$, $\beta > 0$, and $B(\alpha, \beta)$ is a normalizing constant so that the density integrates to 1. When parameterized in this way the mean and variance of the distribution are:

$$\mathrm{E}[X] = \frac{\alpha}{\alpha+\beta}$$

$$\mathrm{V}[X] = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$$

For the purpose of beta regression one isolates the mean of the distribution as a parameter, reparameterizing as follows:

$$\alpha \to \theta\tau, \quad \beta \to (1-\theta)\tau,$$

In this case the mean and variance may be expressed as:

$$\mathrm{E}[X] = \theta \tag{1}$$

$$\mathrm{V}[X] = \frac{\theta(1-\theta)}{\tau+1} \tag{2}$$

Likelihood for individual patient scores

In this section, a beta regression model for patient-level data is described. Let:

- $ADAS_{ipk}$ denote the observed ADAS-cog score on the ith occasion in the pth patient in the kth study;
- t_{ipk} denote the time of the observation relative to the randomization time for that patient,
- D_{ipk} denote the dose assigned to the patient at time t_{ipk}, expressed as a multiple of a reference dose (with reference dose varying by drug).

The indices j and d will be used to identify the study arm and drug (if applicable), where necessary using $j(p)$ and $d(p)$ to express these as functions of the patient index p. We use the notation "| patient p" to denote conditioning on all patient-level random effects associated with patient p (given the hierarchical structure of the model, this also implies conditioning on all study-level random effects).

We specify the residual distribution of scores for patient p using a beta distribution (this is the defining feature of beta regression):

$$ADAS_{ipk}/70 \,|\, \text{patient}\, p \sim \mathrm{Beta}\big(\theta_{ipk}\tau_k,\ (1-\theta_{ipk})\tau_k\big). \tag{3}$$

This distribution is parameterized such that:

$$\mathrm{E}[ADAS_{ipk}/70 \,|\, \text{patient } p] = \theta_{ipk} \tag{4}$$

$$\mathrm{V}[ADAS_{ipk}/70 \,|\, \text{patient } p] = \frac{\theta_{ipk}(1-\theta_{ipk})}{\tau_k+1} \tag{5}$$

Thus, θ_{ipk} is the normalized (i.e. divided by 70) conditional expectation for patient p. This conditional expectation is then related to predictor terms (sometimes referred to as the "systematic component" of the model [1]) via a link function g [3] that maps the interval (0–1) to the entire real line. We adopt a systematic component for our model that is qualitatively similar to that of Ito et al. [9], including terms for an intercept η_{pk}, a slope with respect to time α_{pk}, a placebo effect $E_{\mathrm{PBO}}(t_{ipk})$, (for patients receiving either placebo or active intervention, this is a nonlinear function with respect to time; for patients in the ADNI AD cohort, this is set to zero), and $E_{\mathrm{DRG}}(t_{ipk}, D_{ipk})$ is the effect of drug, where applicable (nonlinear with respect to both time and dose):

$$g(\theta_{ipk}) = \eta_{pk} + \alpha_{pk}t_{ipk} + E_{\mathrm{PBO}}(t_{ipk}) + E_{\mathrm{DRG}}(t_{ipk}, D_{ipk}) \tag{6}$$

Borrowing again from the previous work of of Ito et al. [9], we model the placebo and drug effects as:

$$E_{\mathrm{PBO},ipk} = \beta\big(e^{-k_{el}t_{ipk}} - e^{-k_{eq}t_{ipk}}\big) \tag{7}$$

$$E_{\mathrm{DRG},ipk} = (D_{ipk})^{\gamma_{d(p)}}\frac{E_{\Delta,d(p)}t_{ipk}}{ET_{50,d(p)} + t_{ipk}}. \tag{8}$$

We have referred to our approach as only "qualitatively similar" to that of Ito et al. because the parameter interpretations and implied functional forms on the original scale depend on the choice of link function g. For example, for link functions other than the identity, the additivity of terms within the systematic component does not imply additivity of effects on the original scale, the use of the linear term $\alpha\, t$ to represent natural progression does not imply a linear natural progression on the original scale, the use of a Bateman function for the placebo term does imply that the incremental effect of placebo on the original scale follows an exact Bateman functional form, and so forth. Nonetheless, the link functions of interest are

generally well approximated by linear transformations over limited ranges of the response variable. Thus, when the individual terms contribute only modest increments to the total systematic component (as is the case for all placebo and drug effect terms of interest, as well as disease progression terms over short time frames, e.g. 0–6 months), their incremental effects on the original scale are approximately additive and have longitudinal shapes that are qualitatively similar to the specified functional form.

A typical choice of link function in this context is the standard logit transformation, corresponding to the standard logistic model:

$$g(\theta) = \log\left[\frac{\theta_{ipk}}{(1-\theta_{ipk})}\right] \tag{9}$$

However, as suggested by Samtani et al. [23], generalizations of the standard logistic model may be considered in recognition of potential asymmetries in the progression. In principal, other link functions mapping from the unit interval to the real line (e.g. the probit link, or any inverse cumulative distribution function) could also be considered. For present purposes we have focused on a generalization that may be achieved with the following link function described by Samtani et al:

$$g(\theta) = \log\left[\left(\frac{\theta_{ipk}^{\xi}}{(1-\theta_{ipk}^{\xi})}\right)^{1/\xi}\right] \tag{10}$$

This is one of a number of potential generalizations of the logit transformation that have been proposed [23–25]. While this generalization was not ultimately selected by Samtani et al. the availability of this model in closed form allows for easier generalization to cases where placebo effects and symptomatic drug effects must be accounted for. Moreover, Samtani et al. report that this model and their final selected model were nearly indistinguishable with respect to the Akaike information criterion [15], when applied to the ADNI data set.

The correspondence between our model specification and that of Samtani et al. may be seen by setting both $E_{\mathrm{PBO}}(t_{ipk})$ and $E_{\mathrm{DRG}}(t_{ipk}, D_{ipk})$ to zero in Eq. 6, rewriting θ as a function of time (rather than as a parameter indexed by visit), and applying the inverse of the generalized logit transformation to both sides of the equation, resulting in:

$$\theta_{pk}(t) = (1+\exp(-\xi(\eta_{pk}+\alpha_{pk}t)))^{-1/\xi}. \tag{11}$$

This latter equation may be specified in differential form as:

$$\theta_{pk}(0) = (1+\exp(-\xi\eta_{pk}))^{-1/\xi} \quad \text{(initial condition)} \tag{12}$$

$$\theta'_{pk}(t) = \alpha_{pk}\theta_{pk}(t)[1-\theta_{pk}^{\xi}(t)] \tag{13}$$

This differential equation may be interpreted as a statement that rates of progression are functions of current states (θ_{pk}) and patient-specific rate parameters (α_{pk}).

Patient-level variation in both intercepts and slopes is achieved with patient-level random effects:

$$\eta_{pk}|\ \text{study}\ k \sim \mathrm{N}\left(\mu_{\eta,k},\ \sigma^2_{\eta,k}\right) \tag{14}$$

$$\alpha_{pk}|\ \text{study}\ k \sim \mathrm{N}\left(\mu_{\alpha,k},\ \sigma^2_{\alpha,k}\right) \tag{15}$$

Study-level variation beyond that explained by covariates and inter-subject variation is accommodated by introducing study-level random effects:

$$\mu_{\eta,k} \sim \mathrm{N}\left(v_{\eta},\ \psi^2_{\eta}\right) \tag{16}$$

$$\mu_{\alpha,k} \sim \mathrm{N}\left(v_{\alpha},\ \psi^2_{\alpha}\right) \tag{17}$$

It is important to recognize that inter-subject variance components $\sigma^2_{\eta,\,k}$ and $\sigma^2_{\alpha,\,k}$, rather than reflecting variation in a natural population, are largely determined by the stringency of inclusion / exclusion criteria, which vary by study. Additionally, residual variances can be expected to vary by study (for example, due to different levels of investigator experience and training). It is therefore desirable to allow for potential inter-study heterogeneity in all three intra-study variance components. We implement this heterogeneity as follows:

$$\tau_k \sim \mathrm{Gamma}(\kappa_{\epsilon}, \kappa_{\epsilon}\phi^2_{\epsilon}) \tag{18}$$

$$1/\sigma^2_{\eta,k} \sim \mathrm{Gamma}(\kappa_{\eta}, \kappa_{\eta}\phi^2_{\eta}) \tag{19}$$

$$1/\sigma^2_{\alpha,k} \sim \mathrm{Gamma}(\kappa_{\alpha}, \kappa_{\alpha}\phi^2_{\alpha}) \tag{20}$$

These distributions are parameterized to have means $1/\phi^2_{\epsilon}, 1/\phi^2_{\eta}, 1/\phi^2_{\alpha}$ respectively and coefficients of variation $1/\sqrt{\kappa_{\epsilon}}, 1/\sqrt{\kappa_{\eta}}, 1/\sqrt{\kappa_{\alpha}}$ respectively.

Baseline MMSE (a cognitive assessment used for screening and staging in AD) was employed as a covariate on the $\mu_{\eta,\,k}$ intercept terms, since many predictive inferences of interest to clinical trialists are conditional on MMSE entry criteria for a trial (the MMSE is designed as a screening tool and is almost universally incorporated in inclusion / exclusion criteria, while the ADAS-cog itself is rarely used for this purpose). Moreover, this was the *only* covariate on the intercept terms, since baseline MMSE values are highly correlated with baseline ADAS-cog values [10], and the former is therefore largely sufficient to explain variation in the latter.

All of the covariates that were consistently available across most data sources - baseline MMSE, age, APOE genotype (number of "4" alleles), and gender - were included as covariates on the slope ($\mu_{\alpha,\,k}$) terms, consistent

with the "full covariate model" approach. The use of the first three of these covariates on rate of disease progression is also consistent with the findings of Ito et al. [10] and Samtani et al. [23].

Covariate effects were implemented using the following substitutions in Eqs. 14 and 15 (we use "bMMSE" as notation to refer to baseline MMSE):

$$\mu_{\eta,k} \rightarrow \mu_{\eta,k} + \lambda_{\eta,bMMSE}(bMMSE_{pk} - 21) \tag{21}$$

$$\begin{aligned}\mu_{\alpha,k} \rightarrow\ & \mu_{\alpha,k} + \lambda_{\alpha,bMMSE}(bMMSE_{pk} - 21) \\ & + \lambda_{\alpha,bAge}(bAge_{pk} - 75) \\ & + \lambda_{\alpha,Apo1} I(ApoE_{pk} = 1) + \lambda_{\alpha,Apo2} I(ApoE_{pk} = 2) \\ & + \lambda_{\alpha,Gender} I(Gender_{pk} = Female)\end{aligned} \tag{22}$$

Operational likelihood for summary statistics

Following the approach of Gillespie et al. [7], we model the summary-level data by directly specifying likelihoods based on approximate sampling distributions. Since the model for individual scores is nonlinear, the exact sampling distributions for sample means are not available in analytical form. However, we invoke the approximate linearity of the logit function over the range of primary interest (see Appendix) to derive the approximate distributions, and employ these approximations as our "operational likelihoods". Specifically, using $\overline{ADAS}_{ijk}$ to denote the reported sample mean for arm j in study k, and letting "| arm j" indicate conditioning on all patient-level random effects for all patients in arm j (and, per the model hierarchy, also conditioning on all study-level random effects associated with study k), the individual patient model implies that:

$$\mathrm{E}[\overline{ADAS}_{ijk}/70 |\ \text{arm } j] = \overline{\theta}_{ijk} \equiv \frac{1}{n_{jk}} \sum_{p:j(p)=j} \theta_{ipk}$$

$$\mathrm{V}[\overline{ADAS}_{ijk}/70 \,|\, \text{arm } j] = \left(\frac{1}{n_{jk}^2}\right) \sum_{p:j(p)=j} \mathrm{V}[ADAS_{ipk} |\ \text{patient } p]$$

In addition to knowing the first and second moments of the sample mean, we know that it, like the individual scores, is constrained to the range between zero and seventy. The following Beta distribution therefore, while not an exact likelihood for the normalized sample mean, nonetheless has the basic desired distributional properties and may serve as our "operational" conditional likelihood:

$$\begin{aligned}&\overline{ADAS}_{ijk}/70 \,|\, \text{arm } j \sim \mathrm{Beta}\big(\overline{\theta}_{ijk}(n_{jk}\tau_k + n_{jk} - 1), \\ &(1 - \overline{\theta}_{ijk})(n_{jk}\tau_k + n_{jk} - 1)).\end{aligned}$$

We note that the factor of n_{jk} in the preceding expression provides appropriate weighting of residuals according to sample size, and no other weighting of the conditional residual distribution is required (though additional weighting according to sample size is introduced via the random effects below).

As discussed above, the exact distribution of $\overline{\theta}_{ijk}$ is not analytically available. However, invoking the approximate linearity of the logit transformation over the range of interest, we have:

$$\begin{aligned}&\mathrm{logit}[\overline{\theta}_{ijk}] \approx \overline{\mathrm{logit}[\theta]}_{ijk} \text{ where,} \\ &\overline{\mathrm{logit}[\theta]}_{ijk} \equiv \overline{\alpha}_{jk} t_{ijk} + \overline{\eta}_{\mathrm{intercept},jk} + E_{\mathrm{placebo},ijk} + E_{\mathrm{drug},ijk} \\ &\overline{\alpha}_{jk} \equiv \frac{1}{n_{jk}} \sum_{p:j(p)=j} \alpha_{pk} \\ &\overline{\eta}_{\mathrm{intercept},jk} \equiv \frac{1}{n_{jk}} \sum_{p:j(p)=j} \eta_{\mathrm{intercept},pk}.\end{aligned}$$

The full patient level model would then imply that:

$$\begin{aligned}&\overline{\eta}_{jk} |\ \text{study } k \sim \mathrm{N}(\mu_{\eta,k},\ \sigma_\eta^2/n_{jk}) \\ &\overline{\alpha}_{jk} |\ \text{study } k \sim \mathrm{N}(\mu_{\alpha,k},\ \sigma_\alpha^2/n_{jk})\end{aligned}$$

Missing data mechanism(s) to account for dropout

Since the model attempts to characterize the complete data distribution, summary statistics based on direct simulation from the model are not expected to correctly mimic the behavior of real summary statistics, the latter being computed using incomplete data [4, 6]. Even in the unlikely scenario that the true missing data mechanism (MDM) is missing completely at random (MCAR), the real summary means will be based on fewer observations than their simulated counterparts, and the latter will therefore have standard errors that are unrealistically low. Accordingly, for the purpose of model validation, we incorporate a MDM in our simulations. Parameters for the MDM model were estimated independently from the AD progression parameters (i.e. this was a separate model fitting exercise).

A realistic MDM would be fairly complex and correspondingly would require substantial justification. In order to be concise, we proceed for present purposes with a relatively simple MAR MDM that assumes missingness to be a function of baseline MMSE, age, and study (this last as a random effect). More precisely, we proceed as if missing values occur as the result of attrition, and assuming the times-to-attrition may be described as:

$$T_{pk} \sim \mathrm{Weibull}(\alpha, h_{pk}), \tag{23}$$

where the log of the subject-specific frailty h_{pk} is modeled as:

$$\begin{aligned}\log(h_{pk}) =& \beta_{\mathrm{STUDY},k} + \beta_1(bMMSE_{pk} - 21) \\ & + \beta_2(bAge_{pk} - 75))\end{aligned} \tag{24}$$

The use of baseline MMSE and baseline age as covariates on the dropout hazard borrows from the findings of Faltaos [26].

Note that the proposed model for the MDM was employed only for the purpose of model validation. For the purpose of model fitting, we assume the less restrictive conditions that are required for ignorability of the missing data mechanism, implying that posterior distribution for parameters describing the complete data distribution may be computed using only the observed response and covariates [4, 5]. With respect to patient-level data, the assumptions associated with (approximate) ignorability are arguably moderately realistic. For example, data may be considered missing at random (MAR) with respect to our model even if an individual's missing data status at a particular time point is a function of baseline MMSE, the observed ADAS-cog scores at previous time points, and the treatment assignment (since all of these terms are in the model). With respect to summary-level data the potential biases associated with missing data are not as easily dismissed, however, most of the summary-level data is associated with studies of shorter duration, for which missing data rates are generally low.

Model for missing covariate values

The baseline age, ApoE4 genotype, and gender variables all had missing values for at least some records in the database. Ignoring records with missing covariate values was not considered an acceptable solution, and exclusion of key covariates such as age and ApoE4 genotype was considered highly undesirable. To this end, the model was extended with explicit likelihoods for covariate variables. Given the Bayesian implementation of the model, this approach amounts to averaging over the predictive distribution of missing covariate values. Consequently, response values associated with missing covariates are retained in the analysis, while avoiding biases associated with single imputation methods [14].

The "general location model" [14, 18] was used to specify a joint likelihood for the covariate variables, in recognition of potential stochastic dependencies between the variables. Basic sampling theory was applied to derive the distribution of sample means and proportions under the general location model, and these sampling distributions were specified as the likelihoods for covariates in the meta-data.

Priors

Qualitatively, our intent was to implement priors that constrained parameters to a modest superset of clinically plausible values (i.e. either excluding or rendering unlikely only highly implausible values) and to be otherwise noninformative, consistent with the notion of vague priors [12]. Accordingly, ranges based on clinical plausibility (specified on the original scale) were converted to parametric constraints using the approximate relationships described in the Appendix. In general, the approach was to use Uniform priors within these ranges, although non-Uniform distributions for certain parameters were employed for mathematical and/or computational convenience. The preference for Uniform priors was primarily based on the relative ease with which Uniform priors can be interpreted by scientists from different backgrounds seeking to understand the model. A potential disadvantage of Uniform priors is their lack of support for some parameter values that may technically be plausible. However, density plots of the posterior are generally sufficient to diagnose this situation (i.e. if the posterior mass is concentrated near one of the boundaries of the prior support, the usual implication is that the prior is informative to an undesirable degree). The converse problem, that of offering too much support for parameter values that are implausible and yet cannot be ruled out by the data, may be addressed via sensitivity analyses. Additionally, it is often relatively easy to diagnose the situation where the data provide no information with which to estimate a parameter, since a Uniform posterior that is identical to the prior is easily recognized (indeed, this exact situation arose for the ET_{50} of rivastigmine, alerting us to a lack of sufficient longitudinal data with which to estimate this parameter).

The specific priors on the parametric scale are provided in Table 1.

Model Fitting

Reparameterizations

In order to reduce autocorrelation during Markov Chain Monte Carlo (MCMC) sampling, the following reparameterizations were employed:

$$(\beta, k_{eq}, k_{el}) \longrightarrow \left(-\beta(1/k_{el} - 1/k_{eq}), k_{eq} - k_{el}, k_{el}\right)$$

$$(E_\Delta, ET_{50}) \longrightarrow \left(\frac{(12)E_\Delta}{ET_{50} + 12}, ET_{50}\right)$$

The term $(12)E_\Delta/ET_{50} + 12$ represents the treatment effect at 12 weeks, which is a more directly estimable quantity than the treatment effect at $t = \infty$ The reference time of 12 weeks was selected because the majority of the analyzed trials included assessments at 12 weeks or later.

Computation

Posterior distributions were approximated by Markov Chain Monte Carlo (MCMC) sampling, specifically by

Table 1 Priors and related quantities of interest

Related quantity of interest	Prior (parametric scale)
Population average	
Intercept (points)	$v_\eta \sim$ U (−2.5, 0)
Slope (points per year)	$v_\alpha \sim$ U (0, 0.05)
Covariate adjustments	
On intercepts, ADAS-cog points	
Per point MMSE	$\lambda_{\eta,\ bMMSE} \sim$ (− 0.1, 1)
For slopes, ADAS-cog points per year	
Per point MMSE	$\lambda_{\alpha,bMMSE} \sim$ N (0, 0.01)
Per year of age	$\lambda_{\alpha,\ bAge} \sim$ N (0, 0.01)
per APOE4 heterozygote (vs non-carrier)	$\lambda_{\alpha,\ Apo1} \sim$ N (0, 0.01)
Per APOE4 homozygote (vs non-carrier)	$\lambda_{\alpha,\ Apo2} \sim$ N (0, 0.01)
Per female (vs. male)	$\lambda_{\alpha,Gender} \sim$ N (0, 0.01)
Placebo (incremental) effect	
Area under curve (point-weeks)	$\int E_{\mathrm{PBO}} \sim \mathrm{U}(0, 10)$
Constant for elimination (weeks^{-1})	$k_{el} \sim$ U (0.0001, 0.25)
Constant for onset of placebo effect (weeks^{-1})	$(k_{eq} - k_{el}) \sim$ U (0, 4)
Drug effects (same prior for all AChE inhibitors)	
Difference from placebo at reference dose at 12 weeks (points)	$E_{\mathrm{DRG}}(12, D^*) \sim$ U (0, 0.3)
Time to 50 % of maximum drug effect (weeks)	$ET_{50} \sim$ U (0, 100)
Shape of dose response	$\gamma \sim$ U (0.01, 3)
Inter-study SD	
Of intercepts (points)	$\psi_\eta \sim$ U (0, 1)
Of slopes (points per year)	$\psi_\alpha \sim$ U (0, 0.01)
Cross-study (population) average	
Of inter-individual SD of intercepts (points)	$_\eta,\ \phi_\eta^{\mathrm{M}} \sim$ U (0, 2)
Of inter-individual SD of intercepts (points per year)	$\phi_\alpha,\ \phi_\alpha^{\mathrm{M}} \sim$ U (0, 0.03)
Of residual standard deviation (points)	$\phi_\epsilon,\ \phi_\epsilon^{\mathrm{M}} \sim \mathrm{U}(0, 2)$
Inter-study variance heterogeneity ($1/\sqrt{\kappa_\cdot} = \mathrm{CV}\%$)	
Of inter-individual precisions of intercepts	$\kappa_\eta,\ \kappa_\eta^{\mathrm{M}} \sim$ Pareto (1, 0)
Of inter-individual precisions of slopes	$\kappa_\alpha,\ \kappa_\alpha^{\mathrm{M}} \sim$ Pareto (1, 0)
Of residual precision	$\kappa_\epsilon,\ \kappa_\epsilon^{\mathrm{M}} \sim \mathrm{Pareto}(1, 0)$

Details of conversion between the parametric scale to the original (ADAS-cog) are provided in the Appendix

Gibbs sampling as implemented in WinBUGS version 1.4.3 [16]. For each evaluated model, four distinct MCMC chains were generated, with initial values for most parameters randomly drawn from the prior. In general this resulted in a desirable over-dispersion of initial values as compared to the posterior. Each chain had a total length of 50,000 iterations, of which the first 45,000 "burn-in" samples were discarded from analysis and every 25th of the remaining iterations were kept ("thinning"). Data management, model validation, and model summary were carried out in the R language [19]. To minimize computational time, the MCMC chains were run in PARALLEL. To avoid pseudo-random number generation overlap between chains, the R package, L'Ecuyer et al. [13] was used to assure inter-chain independence. Supporting code is available through http://www.opendiseasemodels.org.

Model evaluation

Model evaluation was broadly comprised of convergence diagnostics, internal validation to assess goodness of fit, and external validation to assess predictive validity. Standard MCMC convergence diagnostics were used including sampling history plots, posterior density estimates, and Gelman-Rubin convergence diagnostics. Internal validation focused primarily on posterior predictive checks based on both study-specific predictions (conditional on study-specific random effect estimates) and marginal predictions (conditional only on covariate values). In addition, external validation proceeded by generating predictions based only on covariate values for a study that had been withheld from the modeling analysts during the model-fitting stage and comparing these predictions to observed values. The

external validation study was randomly selected from a subset of CAMD studies that had sufficient duration (>52 weeks) and size (>1,000 patients) and contained patients with and without background therapy in the control arm.

Model simulation

As is often the case for nonlinear models, many quantities of interest (most notably population intercepts and slopes) were not directly available as parameter estimates. In order to determine the model implications with respect to these quantities, populations of 1,000 patients were simulated for various covariate settings of interest and summary statistics of interest were computed for these simulated populations. Average rates of progression were estimated by simulating populations in the absence of placebo or drug effects and computing the average change from baseline to 52 weeks for the simulated individuals. (Since the model implies a nonlinear progression, this one year average change from baseline is only an approximation to the instantaneous rate of progression at any given time during the first year.)

Results

Data summary

Literature data

A total of 73 studies were collected from the literature, representing 27,895 patients, of which 17,235 patients were in arms used in the analysis (data from control arms other than donepezil, rivastigmine, and galantamine were not included), similar to those described in Ito et al. [9]. A brief summary of the characteristics of the trials collected in the literature database are available on request. Changes in the control arm data demonstrated a "hockey stick" shape, typical in most AD trials (Fig. 1). Following an initial control arm improvement, patients return to a normal progression of disease, which over the course of one to two years, often appears linear, as evidenced by the linearity of the locally weighted scatterplot smooth (loess). Drug treatment arms appear to depart from the normal progression of control arms, but then return to the normal progression, maintaining an offset.

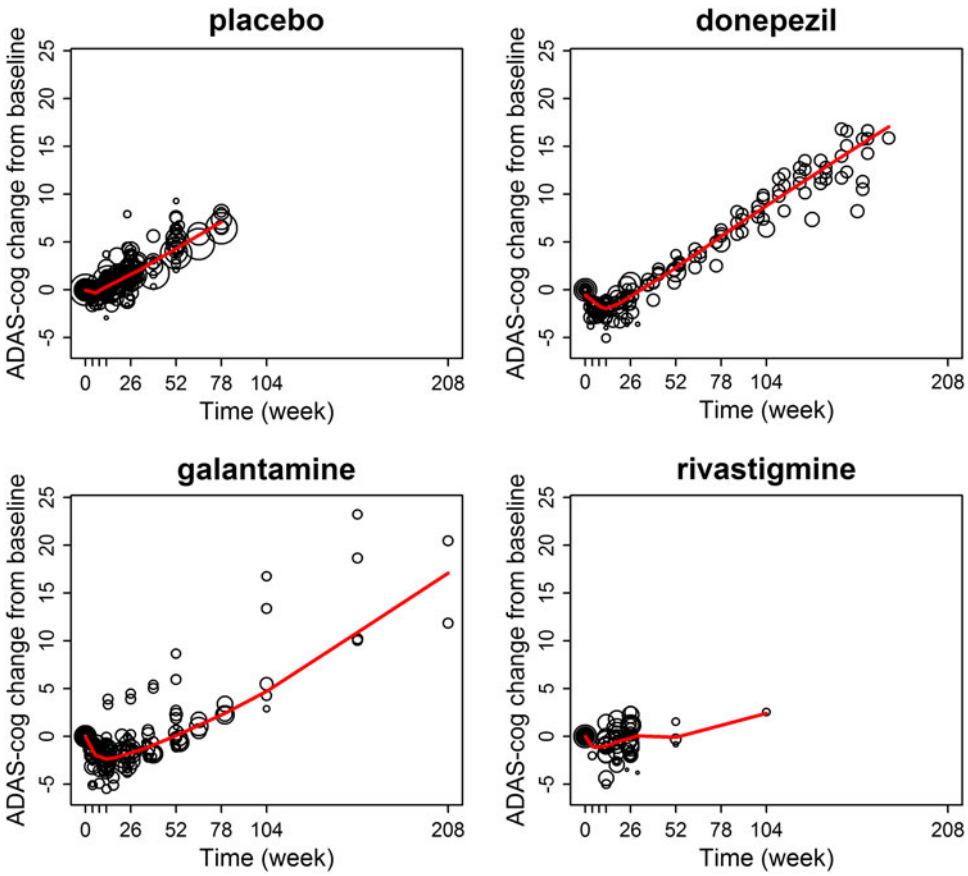

Fig. 1 Observed mean ADAS-cog change from baseline over time by compound from the literature data. *Loess line* is in *red*

CAMD

Table 2 describes the nine studies used for data analysis from the CAMD database, available at the time database development work was initiated in September of 2010. Studies included in the CAMD database after this time were not included. The studies included in the CAMD database consist of all control arm data from all CAMD member-sponsor AD trials in mild and moderate AD that were supplied to CAMD from these studies (both placebo and background therapy arms). Since CAMD focuses on sharing of pre-competitive information, drug treatment arms were not available in the database.

Basic demographics data were similar across the studies in the CAMD database (Table 2). Mean baseline MMSE scores ranged from 19.4 to 21.2 across the studies, with mean age ranging from 72.4 to 75.0 years. ApoE4 status (% e4 carriers, defined as patients with at least one copy of the e4 allele) was also similar for those studies where this information was available. ADAS-cog scores are plotted by study in Fig. 2, while changes from baseline can be seen in Fig. 3.

ADNI

The available dataset contained 817 patients consisting of 229 normal (NL), 402 MCI and 186 AD patients, but only the AD patient data were used for the analysis Average age in the AD cohort was 75.3 ± 7.6, with 47.3% Female. Baseline ADAS-cog and MMSE were 18.7 ± 6.3 and 23.3 ± 2.0 respectively. 33.9 % were ApoE4 carriers and 66.1 were non-carriers. The majority (93.5 %) were white.

Observed longitudinal ADAS-cog data for all ADNI groups are visualized in Fig. 4, showing the clear "floor" effects in the normal elderly.

Final model

The final model was as specified in the Methodology section, with two noted exceptions.

1. Convergence diagnostics suggested that the generalized logistic asymmetry parameter ξ was not well-estimated. Fixing the value of at 1 (resulting in the standard, symmetrical, logistic function) reduced the deviance information criteria (DIC) from −14,363 to −15,115, and this simplification was therefore employed in the final model. Interestingly, the retention of ξ in the model appeared to be advantageous when the literature summary data were removed from the estimation, as in this case the use of the generalized logistic resulted in a decrease of the DIC from −13740.26 to −14137.44. Moreover, the posterior median for ξ in this case was 3.76, corresponding to a sigmoidal inflection point at ADAS-cog = 46 points, substantively very similar to the estimated inflection point at ADAS-cog = 42 points reported by Samtani et al. [23].
2. A graphical review of post-hoc (study-specific) estimates for inter-subject and residual variances revealed substantially discordant estimates for the meta-data in comparison to the individual patient data. Theoretically, our weighting of inter-subject and residual variances according to sample sizes would be expected

Table 2 Studies included in the CAMD database for model development work

Study	1000	1009	1013	1014	1055	1056	1057	1058	1101	Total
Duration (weeks)	12	12	78	78	52	54	54	24	78	–
N	66	164	707	639	140	484	492	162	325	3179
Age (years)	73.7 (8.63)	74.2 (6.36)	74.2 (8.06)	75.0 (8.42)	73.3 (8.16)	72.9 (8.18)	74.2 (7.95)	72.4 (8.70)	73.1 (8.75)	73.9 (8.21)
Female %	54.5	55.5	50.4	56.2	58.6	56.4	61	59.3	51.1	55.3
APOE status (% e4 carriers[a])	–	46.9	–	–	–	59.3	56.1	48.8	59.6	–
bMMSE	20.5 (3.57)	20.6 (3.82)	20.6 (3.30)	21.2 (3.37)	19.4 (3.92)	19.9 (4.22)	19.4 (4.07)	19.5 (4.18)	20.9 (3.60)	20.3 (3.78)
bADAS-cog	19.9 (7.43)	24.2 (11.6)	23.6 (8.82)	21.2 (8.50)	24.7 (10.2)	24.0 (10.4)	25.3 (10.8)	24.8 (10.0)	22.3 (9.70)	23.4 (9.78)
Years since diagnosis	2.6 (<1–8)	<1 (<1–11)	2 (<1–10)	2 (<1–11)	–	2.5 (<1–20)	2 (<1–10)	1.5 (<1–10)	2.4 (<1–12)	2 (<1–20)
Stable background therapy	Yes	No	Both	Both	No	Yes	Yes	No	Yes	Both

[a] ApoE4 carriers include patients with at least one copy of the e4 allele at the APOE locus

Parentheses indicate standard deviation for age, bMMSE, and bADAS-cog, and indicate range for years since diagnosis

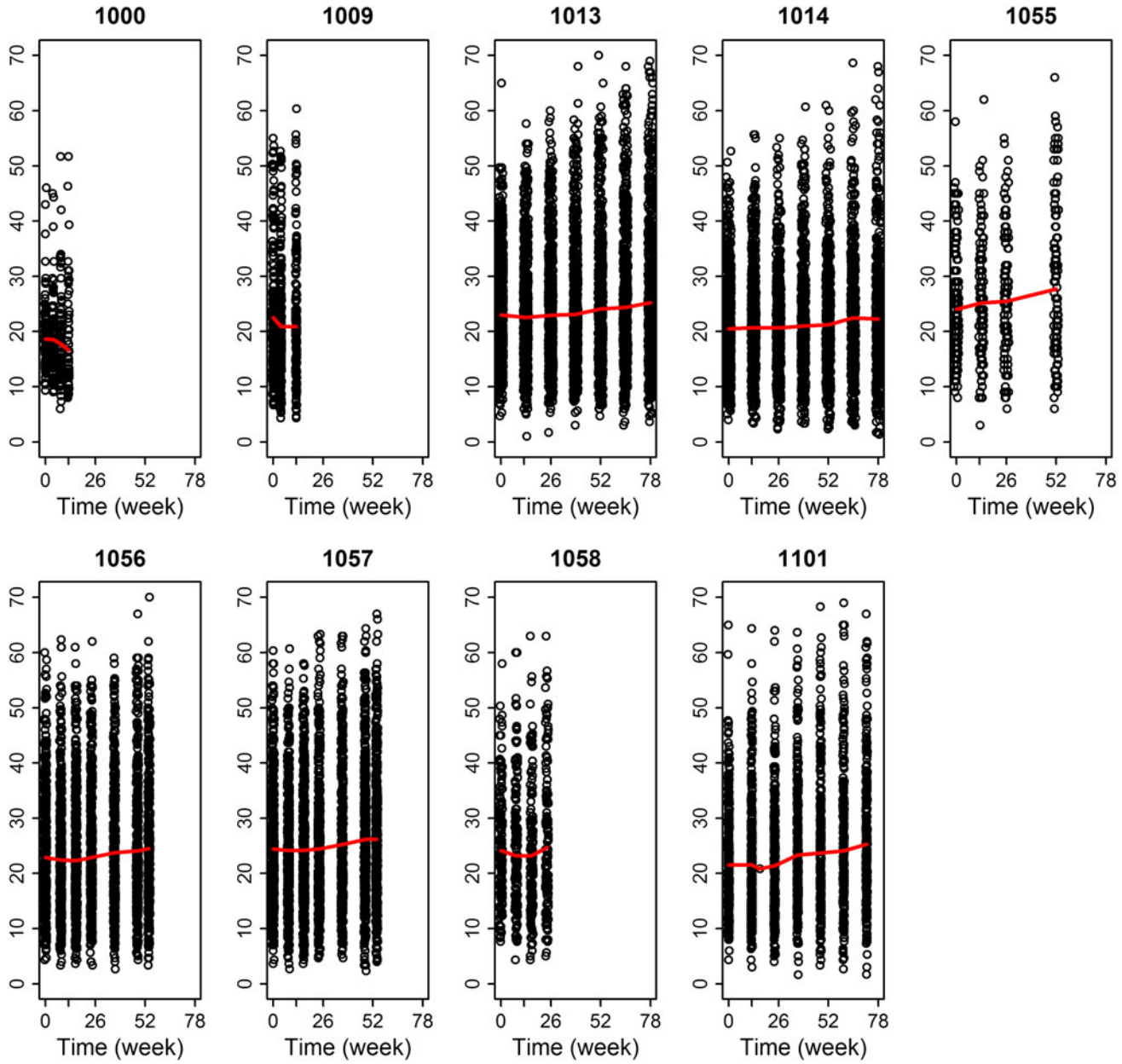

Fig. 2 Observed ADAS-cog scores over time by study in CAMD studies. Locally weighted scatterplot smooth *(loess) line* is shown in *red*

to result in (sample-size normalized) variance components with comparable magnitudes. That these adjustments did not succeed may be at least partly attributable to incomplete sample size information (e.g. in many cases only the number of patients randomized or the number of patients completing the study is reported), and/or due to imputations prior to computations of reported means. This is perhaps cautionary for meta-analyses based only on aggregate data, in which context the present problem would go undiagnosed. In order to remove the influence of the meta-data on the population-level variance components, two distinct sets of variance component parameters (one set for the meta-data and one set for the individual patient data) were employed.

Posterior summaries for all population parameters (i.e. parameters at the highest level of the model hierarchy) are provided in Table 3, which also provides summaries of the approximate posteriors for the estimands on the original scale (the parametric posterior results directly from fitting the model, and the approximate posterior for the estimands on the original scale is obtained by applying the conversions provided in the Appendix).

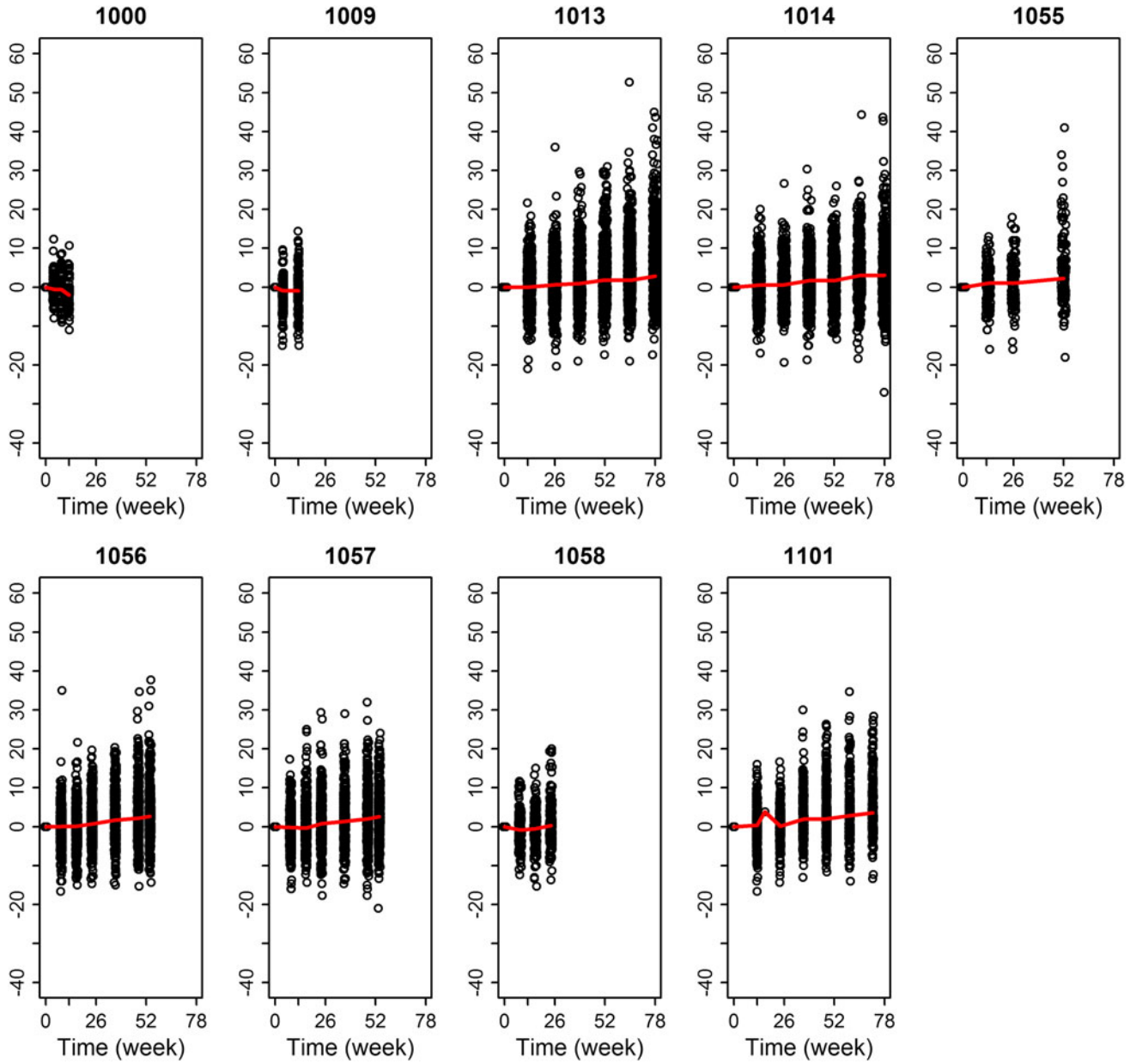

Fig. 3 Observed ADAS-cog changes from baseline over time by study in CAMD studies. Locally weighted scatterplot smooth *loess) line* is shown in *red*

Model Simulation

The model developed provided parameter estimates similar to those described by previous authors [2, 8, 21, 23]. Estimates of baseline ADAS-cog (intercept) from baseline MMSE (Table 4) were consistent with the relationships observed between MMSE and ADAS-cog in the literature, CAMD, and ADNI databases. The parameter estimates observed for covariates of age (Table 5), ApoE4 status, and gender (Table 6) on rate of disease progression on slope were also directionally consistent with previously reported relationships [8, 21, 23]. Specifically, point estimates from the current model suggest that progression rates increase with baseline severity, decrease in baseline age, ApoE4 homozygous carriers progress more rapidly than non-carriers and heterozygotes (though not significant at the 95 % confidence level) , and males progress at a faster rate than females (though not significant at the 95 % level). The effects of age and ApoE4 genotype were simulated and tabulated separately because these two variables are partially confounded, possibly because older patients who were homozygous for ApoE4 were often too far advanced in the disease to be enrolled in the trials. The rates of progression for different baseline severity levels

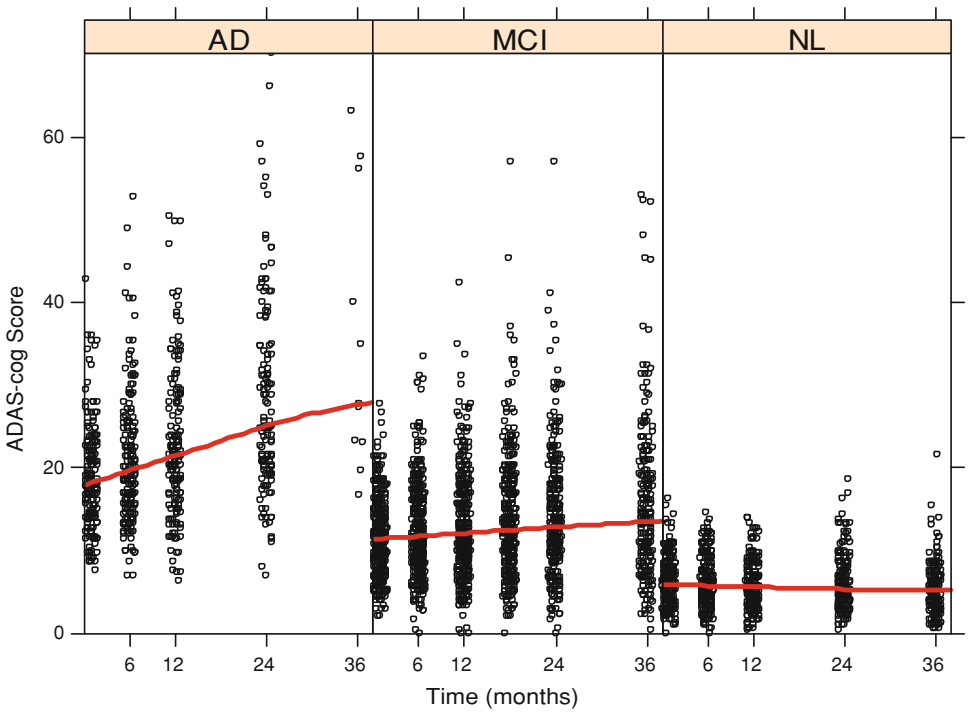

Fig. 4 Longitudinal ADAS-cog by patient population in ADNI study. *Loess line* is in *red*

(Table 4) also align with previously reported values [8, 21].

To visualize the model implications with respect to the dependence of slope and intercept on baseline MMSE, refer to Fig. 5, which shows the time-course of expected ADAS-cog scores for "average individuals" (all random effects set to zero, at a reference covariate setting) from AD subpopulations with varying levels of baseline MMSE. One notes that for the range of baseline MMSE scores that are represented in our primary data sources, the nonlinearities in the expected time-courses are practically negligible. Note this refers only to the expected time-courses for "average individuals" in each MMSE-defined subpopulation. By contrast, nonlinearities are apparent at the individual level, as was seen in posterior predictive checks for individual profiles (not shown here). Figure 5 demonstrates the nonlinearities implied by the model over a longer period of time (Fig. 6).

Model evaluation

Posterior predictive checks

Figure 7 provides unconditional predictive checks (not conditioned on study-level random effects), for studies that were included in the model building from the CAMD and ADNI datasets. The results suggest that an adequate fit was established.

External validation

The response data from the selected CAMD protocol was withheld and blinded from model developers during the model development phase. The final model was then used to generate a predictive distribution for the withheld response data, given the covariate values for that study, in a manner identical to that used for the internal validation "unconditional" predictive checks. The predictive validity of the model was then assessed by graphically comparing the observed data to the model predictions (Fig. 8) to determine if all values fell within the 90 % prediction interval.

Dropout model summary and evaluation

The fitted dropout model exhibited a high degree of agreement with the observed dropout rates, as seen in Fig. 9. The model adequately captures the dropout rate both by baseline MMSE and by age in these two plots.

Discussion

Early models described ADAS-cog as a linear progression as a function of the baseline score, and a progression of the disease as a function of time [20]. The Ito literature model identified that the severity of the disease itself influenced

Table 3 Summary of posterior distribution on parametric scale

Description of estimand	Parameter	Estimated posterior quantile		
		2.5th %ile	Median	97.5th %ile
Population average				
Intercept (points)	v_η	-0.83	-0.80	-0.77
Slope (points per year)	v_α	3.81×10^{-3}	5.07×10^{-3}	6.08×10^{-3}
Covariate adjustments				
Intercepts				
Per MMSE point	$\lambda_{\eta,bMMSE}$	-0.13	-0.13	-0.12
Slopes				
Per MMSE point	$\lambda_{\alpha,bMMSE}$	-7.84×10^{-4}	-6.76×10^{-4}	-5.68×10^{-4}
Per year of age	$\lambda_{\alpha,bAge}$	-1.59×10^{-4}	-1.14×10^{-4}	-6.98×10^{-5}
For ApoE4 heterozygote vs non-carrier	$\lambda_{\alpha,\ Apo1}$	-6.52×10^{-5}	-2.49×10^{-5}	1.96×10^{-5}
For ApoE4 homozygote vs non-carrier	$\lambda_{\alpha,Apo2}$	-1.12×10^{-5}	1.25×10^{-4}	2.60×10^{-4}
For female vs. male	$\lambda_{\alpha,Gender}$	-1.36×10^{-3}	-6.53×10^{-4}	9.92×10^{-5}
Placebo (incremental) effect				
Area under curve (point-weeks)	$\int E_{\mathrm{PBO}}$	3.72	5.30	9.10
Constant of elimination (weeks^{-1})	k_{el}	1.15×10^{-2}	2.75×10^{-2}	4.87×10^{-2}
Constant of onset (weeks^{-1})	k_{eq}	4.91×10^{-2}	8.08×10^{-2}	0.17
Drug (incremental) effect at 12 weeks				
Donepezil 10 mg QD (points)	$E_{\mathrm{don}}(12, 10)$	0.11	0.14	0.17
Galantamine 24 mg QD (points)	$E_{\mathrm{gal}}(12, 24)$	0.13	0.16	0.18
Rivastigmine 6 mg QD (points)	$E_{\mathrm{riv}}(12, 6)$	7.43×10^{-2}	0.13	0.17
Onset of drug effects (ET_{50})				
Donepezil (weeks)	$ET_{50,\ \mathrm{don}}$	0.36	1.62	4.94
Galantamine (weeks)	$ET_{50,\ \mathrm{gal}}$	5.25	8.98	15.48
Rivastigmine (weeks)	$ET_{50,\ \mathrm{riv}}$	0.39	7.90	75.23
Shape of dose response				
Donepezil	γ_{don}	3.35×10^{-2}	0.24	0.59
Galantamine	γ_{gal}	1.95×10^{-2}	0.20	0.55
Rivastigmine	γ_{riv}	1.33×10^{-2}	0.11	0.53
Inter-study SD				
Of intercepts	ψ_η	5.80×10^{-2}	9.54×10^{-2}	0.136
Of slopes	ψ_α	8.39×10^{-4}	1.64×10^{-3}	2.80×10^{-3}
Cross-study (population) average				
Of inter-individual SD of intercepts	$_\eta$	0.38	0.41	0.42
Of inter-individual SD of slopes	$_\alpha$	6.07×10^{-3}	7.78×10^{-3}	9.95×10^{-3}
Of residual SD	ϕ_ϵ	9.97×10^{-2}	0.11	0.11
Between-study variation (CV%)				
In inter-individual precisions of intercepts	$1/\sqrt{\kappa_\eta}$	3.15×10^{-2}	0.14	0.39
In inter-individual precisions of slopes	$1/\sqrt{\kappa_\alpha}$	0.29	0.61	1.18
In residual precision	$1/\sqrt{\kappa_\epsilon}$	5.40×10^{-2}	0.12	0.28

the slope, and thus the slope changed over time (introducing non-linearity). More recent models directly introduced non-linear relationships to describe the course of disease over time. Earlier models also varied with respect to the completeness of components required of a drug-disease-trial model. Of those described, none had all components required that describe existing drug effects, the underlying disease progression, and all pertinent factors be considered in a clinical trial. A drug-disease-trial model that includes all these components would require underlying data that can inform each of the various trial components in the model. For example, natural history data

Table 4 Model predicted expected mean ADAS-cog score intercept by bMMSE

BMMSE	Median	5 % LB	95 % UB
16	32.2	29.6	34.9
21	22.3	20.0	24.8
26	14.2	12.4	16.1

Table 5 Model predicted expected mean change in ADAS-cog score over one year in the absence of a placebo or drug effect, by baseline age

Age	Median	5% LB	95% UB
69	4.92	3.71	6.13
75	4.39	3.51	5.39
80	4.00	2.97	5.17

Table 6 Model predicted expected mean change in ADAS-cog score over one year in the absence of placebo or drug effect, by baseline MMSE, gender, and ApoE4 status

BMMSE	Gender	ApoE4	Median	5 % LB	95 % UB
16	Male	0	7.14	4.48	9.54
16	Male	1	7.07	4.49	9.42
16	Male	2	8.03	5.20	10.40
16	Female	0	6.53	3.73	9.05
16	Female	1	6.52	3.88	9.04
16	Female	2	7.55	4.76	9.78
21	Male	0	4.48	1.99	7.09
21	Male	1	4.43	2.06	6.94
21	Male	2	5.43	2.82	8.06
21	Female	0	3.97	1.42	6.57
21	Female	1	3.97	1.52	6.59
21	Female	2	4.88	2.16	7.17
26	Male	0	1.69	-0.28	4.12
26	Male	1	1.70	-0.33	4.00
26	Male	2	2.39	0.34	4.90
26	Female	0	1.36	-0.61	3.78
26	Female	1	1.35	-0.68	3.71
26	Female	2	2.01	0.02	4.53

inform estimation of underlying disease progression and provide and a rich source of covariates for model building, placebo arm data inform estimation of the magnitude, onset and offset of placebo response in controlled clinical trials as well as the rate of drop-out in the trials, and summary data from the literature support estimates of various drug

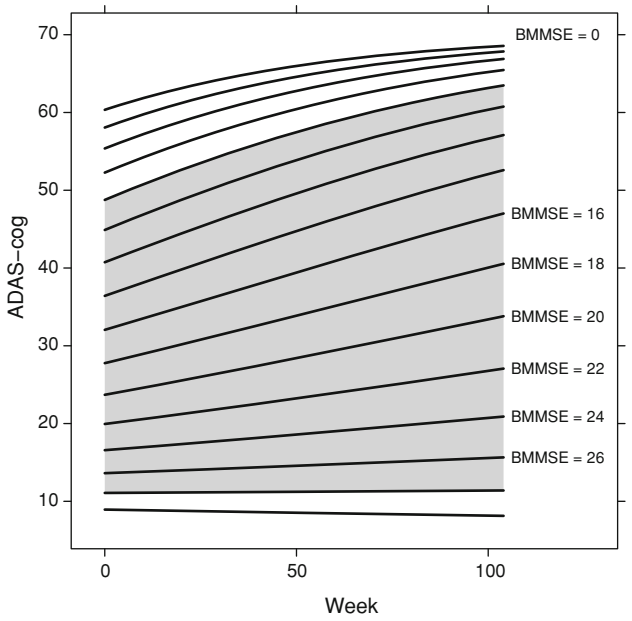

Fig. 5 Predicted two year "natural progression" (i.e. in the absence of any intervention, including placebo) for an "average individual" (i.e. with all random effects set to zero, at reference covariate settings age = 75, male, ApoE4 non-carrier) as a function of baseline MMSE. Predictions well-supported by the data are those in the grey shaded area; predictions outside of this range are extrapolations. All predictions refer to simulated subjects with a diagnosis of probable AD. Thus, to interpret the prediction for baseline MMSE = 30 (for example), one must envision a population of patients who are first diagnosed with probable AD (perhaps based on an initial screening MMSE assessment) and subsequently obtain a baseline MMSE score (perhaps based on a baseline assessment that is distinct from the screening assessment) of 30. To the extent that such subpopulations are unlikely in practice, one may simply view the prediction for MMSE = 30 as a mathematical extrapolation

effects (magnitude, time to onset, and durability). Unfortunately, no single trial could provide all of these elements.

The proposed BR model is both theoretically coherent and practically effective for simultaneously describing multiple aspects of a large and diverse set of data, while maintaining a reasonable degree of parsimony and interpretability. It incorporates key learnings from previous authors, and introduces new enhancements. The model fit, being based on a more comprehensive collection of data than has been used previously, arguably represents our most reliable and accurate estimates to date based on the totality of the available evidence. The current model provides a sufficient basis for simulating a wide variety of clinical trial designs in order to determine their operating characteristics.

BR models have broad utility, perhaps under-recognized in pharmacometrics, for analyzing constrained response data. In general, BR models may be implemented on a variety of software platforms. While the meta-analytic aspect of our analysis requires an extension of typical beta-regression

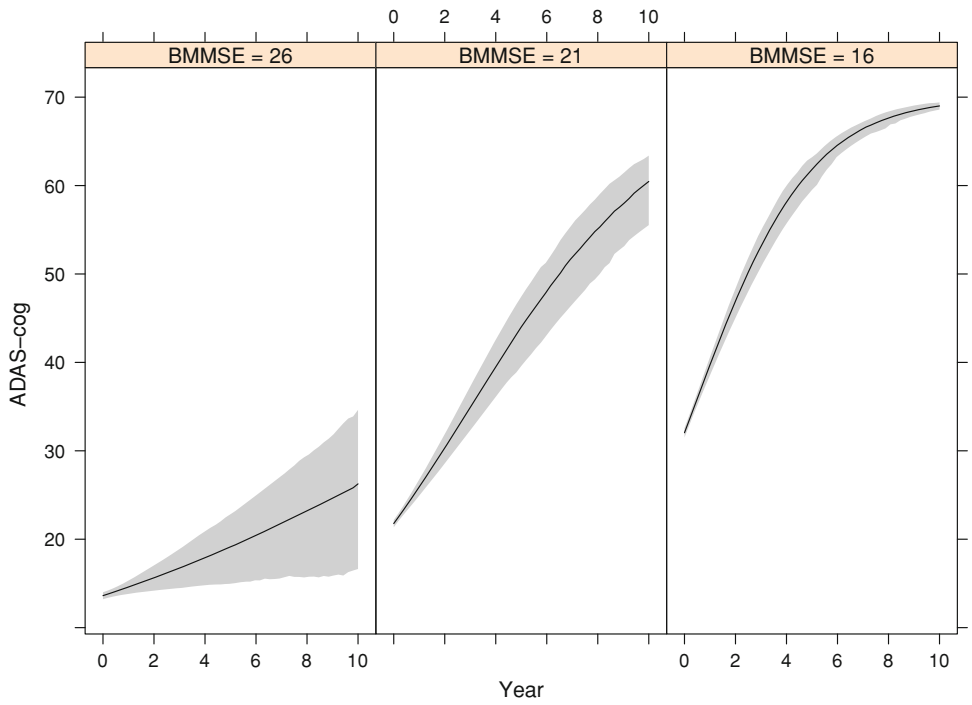

Fig. 6 *Lines* represent posterior median predictions for an "average individual" (i.e. with all random effects set to zero, at reference covariate settings age = 75, male, ApoE4 non-carrier) and grey region represents the corresponding 90 % credible interval for the predictions. Predictions past two years represent extrapolations beyond the extent of the available data, and are intended primarily to show that the mathematical implications of the model are consistent with the expected nonlinear progression of the endpoint

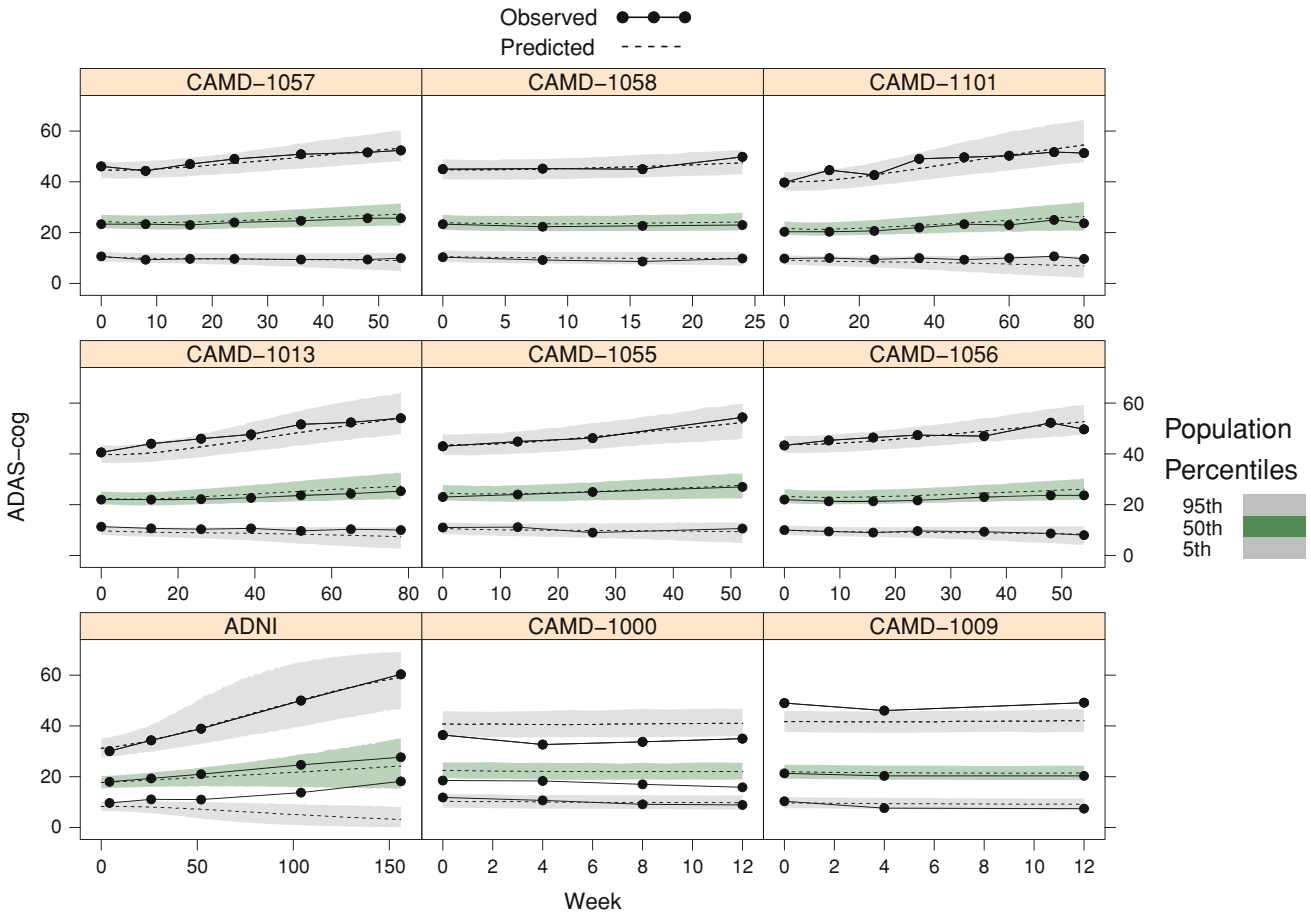

Fig. 7 Plot of unconditional predictive checks for sample population percentiles of ADNI and CAMD studies

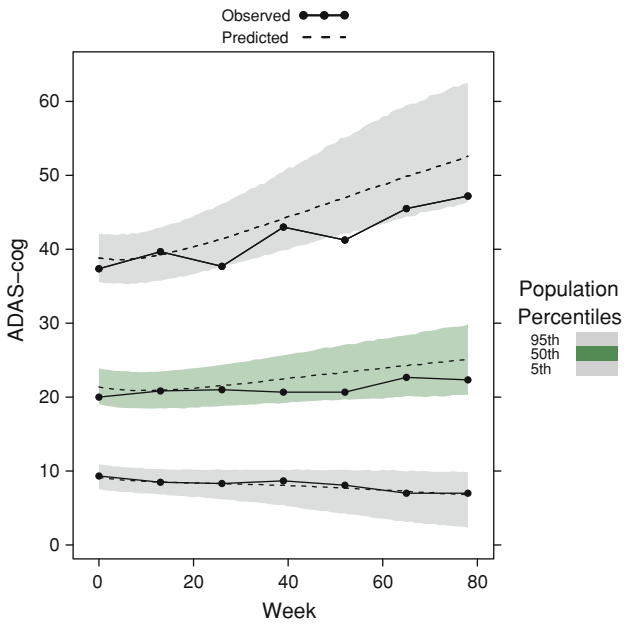

Fig. 8 Plot of model prediction for a study selected (1014) (external validation)

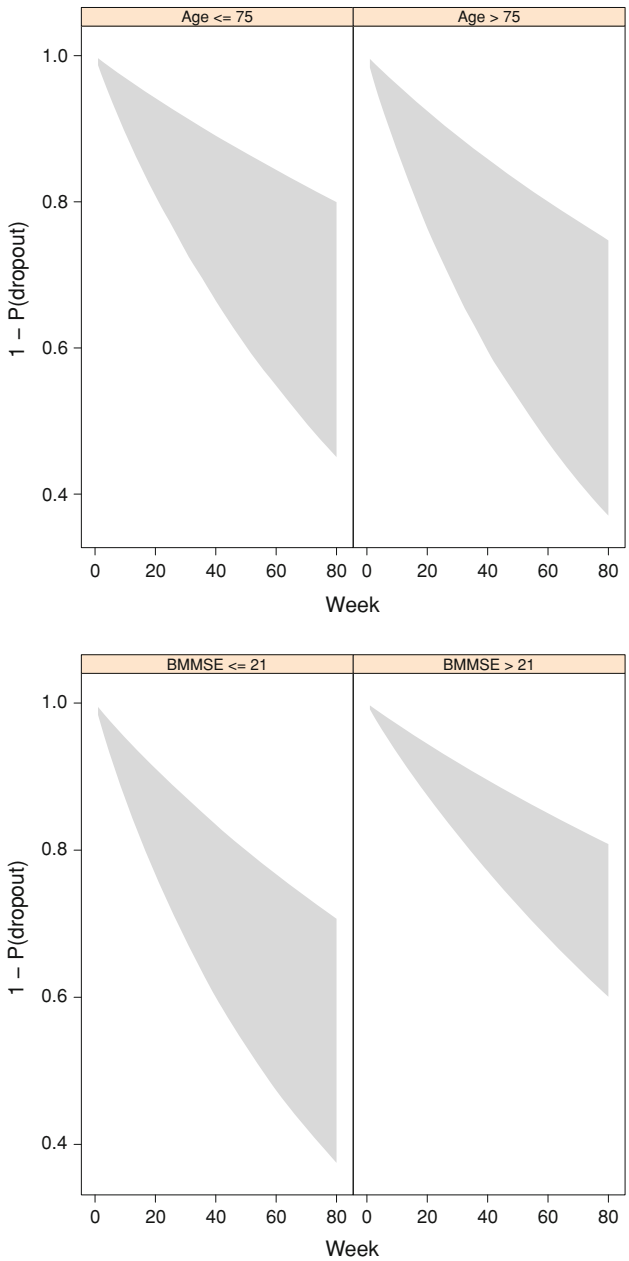

Fig. 9 Plot of probability (dropout) over time by baseline age (*upper panel*) and baseline MMSE (*lower panel*)

machinery and we are uncertain whether our particular hierarchical beta-regression model could be easily implemented outside of WinBUGS, application of our model to simpler patient-level data sets appears to be straightforward in popular modeling software. For example, NONMEM allows for specification of an arbitrary likelihood for residuals, and beta regression generally employs the beta distribution only for the residual likelihood. In R, a flexible class of beta regression models has been implemented in the package betareg (http://cran.r-project.org/web/packages/betareg/vignettes/betareg-ext.pdf).

To our knowledge, this is the first integration of patient level and literature level data for development of a drug-disease-trial tool. The implementation of this type of integrative model provided a complete knowledge management framework, allowing clinical trial design teams to utilize all available knowledge, to simulate myriad trial design scenarios to answer specific questions within a quantitative framework. Conversely, we note certain limitations of this first application of this work. The predictive implications of the model need to be investigated under a wider variety of plausible missing data mechanisms. This is the case for model validation purposes (specifically, for posterior predictive checks) as well as for future clinical trial simulations to evaluate trial designs. Additionally, our current strategy for handling missing covariate values may be unwieldy when considering a broader set of covariates that are missing at higher rates.

The work described here was completed to support the CAMD, whose goal is to provide a pre-competitive space to allow health authorities, industry, and patient advocacy groups to work together to develop tools to facilitate and accelerate drug development in neurodegeneration. In that spirit, we encourage interested parties to visit C-Paths CODR page (http://codr.c-path.org) a unique resource for Critical Path Institute consortia members and qualified researchers to upload and work on valuable scientific data, relevant to biomarkers of drug toxicity, neurodegenerative diseases, and patient-reported outcomes; as well as the Alzheimers Disease Progression site model project http://www.opendiseasemodels.org, through which the combined

data set, all R code and WinBUGS model files, training documentation, and a number of supplementary diagnostic figures and tables may be retrieved. The Open Disease Models website, which links to an associated Googlecode git repository (http://code.google.com/p/alzheimers-disease-progression-model-adascog) provides a platform for interested parties to contribute model criticism, additional code (including edits to existing code if authorized), and supplementary analyses.

Acknowledgements Critical Path Institutes Coalition Against Major Diseases is supported by grant No. U01FD003865 from the United States Food and Drug Administration and by Science Foundation Arizona under Grant No. SRG 0335-08. Data collection and sharing for the ADNI component of our combined data set was funded by the AD Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: Abbott, AstraZeneca AB, Bayer Schering Pharma AG, Bristol-Myers Squibb, Eisai Global Clinical Development, Elan Corporation, Genentech, GE Healthcare, GlaxoSmithKline, Innogenetics, Johnson and Johnson, Eli Lilly and Co., Medpace, Inc., Merck and Co., Inc., Novartis AG, Pfizer Inc, F. Hoffman-La Roche, Schering-Plough, Synarc, Inc., as well as non-profit partners the Alzheimer's Association and Alzheimer's Drug Discovery Foundation, with participation from the U.S. Food and Drug Administration. Private sector contributions to ADNI are facilitated by the Foundation for the National Institutes of Health (http://www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Disease Cooperative Study at the University of California, San Diego. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of California, Los Angeles. This research was also supported by NIH grants P30 AG010129, K01 AG030514, and the Dana Foundation. Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: Abbott; Alzheimers Association; Alzheimers Drug Discovery Foundation; Amorfix Life Sciences Ltd.; AstraZeneca; Bayer HealthCare; BioClinica, Inc.; Biogen Idec Inc.; Bristol-Myers Squibb Company; Eisai Inc.; Elan Pharmaceuticals Inc.; Eli Lilly and Company; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; GE Healthcare; Innogenetics, N.V.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Medpace, Inc.; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Servier; Synarc Inc.; and Takeda Pharmaceutical Company. 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 (http://www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Disease Cooperative Study at the University of California, San Diego. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of California, Los Angeles. This research was also supported by NIH grants P30 AG010129, K01 AG030514, and the Dana Foundation.

Appendix: Linear approximation to the logit transformation

One may obtain linear approximations to the logit and inverse logit transformations using first-order Taylor expansion of $\text{logit}(x)$ around $x = 0.5$, resulting in the

Table 7 Approximate transformations between parametric scale and original scale

Description of estimand	Approximation as function of parameters
Population average	
Intercept (points)[a]	$70 \times 1/(1+\exp(-v_\eta))$
Slope (points per year)[a]	$52 \times 70 \times v_\alpha / 4$
Covariate adjustment per point MMSE	
For intercepts (points ADAS-cog per point MMSE)	$70 \times \lambda_\eta / 4$
For slopes (points ADAS-cog per year per point MMSE)	$52 \times 70 \times \lambda_\alpha / 4$
Placebo (incremental) effect	
Area under curve (point-weeks)	$70 \times \int E_{\text{PBO}}/4$
Constant for elimination (weeks^{-1})	k_{el}
Constant for onset of placebo effect (weeks^{-1})	k_{eq}
Drug effects (same prior for all AChE inhibitors)	
Difference from placebo at reference dose at 12 weeks (points)	$70 \times E_{\text{drug}}(12, D^*)$ /4
Time to 50 % of maximum drug effect (weeks)	ET_{50}
Shape of dose response	γ
Inter-study SD	
Of intercepts (points)	$70 \times \psi_\eta / 4$
Of slopes (points per year)	$70 \times \psi_\alpha / 4$
Cross-study (population) average	

Table 7 continued

Description of estimand	Approximation as function of parameters
Of inter-individual SD of intercepts (points)	$70 \times \phi_\eta$ / 4
Of inter-individual SD of intercepts (points per year)	$52 \times 70 \times \phi_\alpha$ /4
Of residual standard deviation (points)[b]	$70 \times \sqrt{\left(\frac{25}{70}\right) \times \left(1 - \frac{25}{70}\right) \times \frac{1/\tau_\epsilon}{(1+1/\tau_\epsilon)}}$
Between-study variation (CV%)	
In inter-individual precisions of intercepts	$100 \times 1/\sqrt{\kappa_\eta}$
In inter-individual precisions of slopes	$100 \times 1/\sqrt{\kappa_\alpha}$
In residual precision	$100 \times 1/\sqrt{\kappa_\epsilon}$

[a] Population averages for slope and intercept refer to a reference sub-population with MMSE = 21

[b] Use of the Beta distribution for residuals implies that the residual standard deviation varies as a function of expected value; therefore, for purposes of approximate conversion to the original scale, we select a reference expected value of ADAS-cog = 25

approximation $\text{logit}(ADAS/70) \approx -2 + 4 \times (ADAS/70)$, or for the inverse relationship. This linear approximation is not invoked for modeling per se, although the justification of our "operational likelihood" for summary statistics does invoke the existence of a linear approximation. Rather, its primary utility is to enable an approximate interpretation of model parameters on the original scale, this being important for both prior specification and model (posterior) summary.

We tabulate the approximate transformations used for each parameter in Table 7. With the exception of the population average intercept, all conversions rely on the linear approximation described above (in the case of the population average intercept, one may apply the inverse logit transformation directly, since this is a location parameter; similar exact transformations are not possible for scale parameters). Where appropriate, a multiplicative factor of 52 is also used to convert weekly rates to annual rates.

References

1. Agresti A (1990) Categorical data analysis. Wiley, New York
2. Ashford JW, Schmitt FA (2001) Modeling the time-course of alzheimer dementia. Curr Psychiatry Rep 3(1):20–28
3. Ferrari S, Cribari-Neto F (2004) Beta regression for modelling rates and proportions. J Appl Stat 31(7):799–815
4. Gastonguay MR, French JL, Heitjan DF, Rogers JA, Ahn JE, Ravva P (2010) Missing data in model-based pharmacometric applications: points to consider. J Clin Pharmacol 50 (Suppl 1):63S
5. Gelman A, Carlin JB, Stern HS, Rubin DB (2004) Bayesian data analysis. Chapman & Hall/CRC, New York
6. Gelman A, Van MI, Verbeke G, Heitjan DF, Meulders M (2005) Multiple imputation for model checking: completed-data plots with missing and latent data. Biometrics 61(1):74–85
7. Gillespie WR, Rogers JA, Ito K, Gastonguay MR (2009) Population dose-response model for adas-cog scores in patients with alzheimers disease by meta-analysis of a mixture of summary and individual data. In: American Conference on Pharmacometrics. Mashantucket
8. Holford NH, Peace KE (1992) Methodologic aspects of a population pharmacodynamic model for cognitive effects in alzheimer patients treated with tacrine. Proc Natl Acad Sci 89(23): 11,466–11,470 http://www.pnas.org/content/89/23/11466.abstract
9. Ito K, Ahadieh S, Corrigan B, French J, Fullerton T, Tensfeldt T (2010) Alzheimer's disease working group: disease progression meta-analysis model in alzheimer's disease. Alzheimers Dement 6(1):39–53. doi:10.1016/j.jalz.2009.05.665
10. Ito K, Corrigan B, Zhao Q, French J, Miller R, Soares H, Katz E, Nicholas T, Billing B, Anziano R, Fullerton T (2011) Alzheimer's disease neuroimaging initiative: disease progression model for cognitive deterioration from alzheimer's disease neuroimaging initiative database. Alzheimers Dement 7(2):151–160. doi:10.1016/j.jalz.2010.03.018
11. Kieschnick R, McCullough BD (2003) Regression analysis of variates observed on (0, 1): percentages, proportions and fractions. Stat Model 3:192–213
12. Lambert PC, Sutton AJ, Burton PR, Abrams KR, Jones DR (2005) How vague is vague? a simulation study of the impact of the use of vague prior distributions in mcmc using winbugs. Stat Med 24(15):2401–2428. doi:10.1002/sim.2112
13. L'Ecuyer P, Simard R, Chen EJ, Kelton WD (2001) An object-oriented random-number package with many long streams and substreams. Oper Res 50:1073–1075
14. Little RJA, Rubin DB (2002) Statistical analysis with missing data, 2nd edn. Wiley, New York
15. Ludden TM, Beal SL, Sheiner LB (1994) Comparison of the akaike information criterion, the schwarz criterion and the f test as guides to model selection. J Pharmacokinet Biopharm 22(5):431–445
16. Lunn D, Thomas A, Best N, Spiegelhalter D (2000) Winbugs—a bayesian modelling framework: concepts, structure, and extensibility. Stat Comput 10:325–337
17. Mohs RC, Rosen WG, Davis KL (1983) The alzheimer's disease assessment scale: an instrument for assessing treatment efficacy. Psychopharmacol Bull 19(3):448–450
18. Olkin I, Tate RF (1961) Multivariate correlation models with mixed discrete and continuous variables. Ann Math Stat 32(2): 448–465
19. R Development Core Team (2008) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna. http://www.R-project.org. ISBN: 3-900051-07-0
20. Romero K, Corrigan B, Neville J, Kopko S, Cantillon M (2011) Striving for an integrated drug-development process for

neurodegeneration: the coalition against major diseases. Neurodegener Dis Manag 1(5):379–385
21. Romero K, de Mars M, Frank D, Anthony M, Neville J, Kirby L, Smith K, Woosley RL (2009) The coalition against major diseases: developing tools for an integrated drug development process for Alzheimer's and Parkinson's diseases. Clin Pharmacol Ther 86(4):365–367. doi:10.1038/clpt.2009.165
22. Rosen WG, Mohs RC, Davis KL (1984) A new rating scale for alzheimer's disease. Am J Psychiatry 141(11):1356–1364
23. Samtani MN, Farnum M, Lobanov V, Yang E, Raghavan N, DiBernardo A, Narayan V (2012) Alzheimer's disease neuroimaging initiative: an improved model for disease progression in patients from the Alzheimer's disease neuroimaging initiative. J Clin Pharmacol 52(5): 629–644
24. Stukel TA (1988) Generalized logistic models. J Am Stat Assoc 83(42):426–431
25. Tsoularis A, Wallace J (2002) Analysis of logistic growth models. Math Biosci 179(1):21–55
26. Williams-Faltaos D, Ying C, Wang Y, Gobburu J, Zhu H, Quantification of disease progression and drop-out for Alzheimer's disease. Technical report, Food and Drug Administration