EUROPEAN MEDICINES AGENCY
SCIENCE MEDICINES HEALTH

12 July 2013
EMA/CHMP/SAWP/420174/2013
Human Medicines Development and Evaluation

Qualification opinion of a novel data driven model of disease progression and trial evaluation in mild and moderate Alzheimer's disease

Draft agreed by Scientific Advice Working Party	6 June 2013
Adopted by CHMP for release for consultation	27 June 2013[1]
Start of public consultation	19 July 2013[2]
End of consultation (deadline for comments)	27 August 2013[3]

Comments should be provided using this template. The completed comments form should be sent to Qualification@ema.europa.eu

Keywords	Qualification opinion, model of disease progression, mild and moderate Alzheimer's disease

[1] Last day of relevant Committee meeting.
[2] Date of publication on the EMA public website.
[3] Last day of the month concerned.

7 Westferry Circus ● Canary Wharf ● London E14 4HB ● United Kingdom
Telephone +44 (0)20 7418 8400 **Facsimile** +44 (0)20 7523 7040
E-mail info@ema.europa.eu **Website** www.ema.europa.eu
An agency of the European Union

Introduction

On 20 March 2013 the Applicant Critical Path Global Ltd. requested qualification opinion for the proposed Disease Progression and Trial Evaluation Model.

The context of use: “The proposed Disease Progression and Trial Evaluation Model, as defined in this document, is suitable for qualification for use in Drug development as a longitudinal model for describing changes in cognition in patients with mild and moderate AD, and for use in trial designs in mild and moderate AD.”

Dr David Brown was appointed as coordinator. The Qualification Team comprised of: Dr Susan Morgan Mr Rob Hemmings, Dr Ferran Torres, Dr Bertil Jonsson, Dr Monique Wakelkamp, Dr Valentina Mantua, Prof Luca Pani. The Patient representative for the procedure was Mr Jean Georges. The EMA Scientific Administrator for the procedure was Dr Maria Isaac.

The procedure started during the SAWP meeting held on 02 – 04 April 2013.

The Qualification Team meeting took place on 07 May 2013. The discussion meeting with the Applicant took place on 04 June 2013.

During its meeting held on 03 – 06 June 2013, the SAWP agreed on the opinion to be given to the Applicant. During its meeting held on 24 – 27 June 2013, the CHMP adopted the opinion to be given to the Applicant.

The work described in this submission is intended to produce a tool that provides a common quantitative basis for the evaluation of study design and analysis methodologies for clinical studies in the mild and moderate AD population, with cognition as a primary endpoint. To this end, the CAMD effort described here has focused on assessing the progression of the disease as measured by the ADAS-Cog, the most common primary outcome for cognition in this population.

Intended context of use & scope

The context of use

“The proposed disease progression and trial evaluation model, as defined in this document, *is suitable for qualification* for use in drug development as a longitudinal model for describing changes in cognition in patients with mild and moderate AD, and for use in assisting in trial designs in mild and moderate AD.”

The intended scope and use for the drug development tool presented in this application are as follows:

General area: the goal is for this tool to serve as a resource for sponsors designing trials across the Alzheimer’s community. It is intended that sponsors will utilize this simulation tool to provide a quantitative rationale for selection of study design and inclusion criteria. This tool could also be utilized by sponsors and health authorities for comparison of post-hoc analysis results to historical controls (priors) to minimize false positives.

Target population for use: mild and moderate dementia of the Alzheimer’s type.

Intended trial endpoint for use: ADAS-Cog, in trials intended for mild and moderate AD, for study durations of up to two years.

Stage of drug development for use: all clinical stages of AD drug development, including proof of concept, dose-ranging and confirmatory trial designs, and encompassing various types of treatment mechanisms (symptomatic and disease modifying).

Intended applications: potential applications of this tool are an aid for: alternative method for sample size calculations determination of optimal trial durations and measurement times comparison of the sensitivity of competing trial designs to assumptions about the types of expected treatment effects (time to maximal effect, effects that increase or decrease over time), impact of inclusion criteria/disease severity on treatment effect and required trial length determination of the most appropriate data analytic methods for novel trial designs. The model may also be suitable as an informed prior for critical evaluation of retrospective or post-hoc analyses, to minimize the likelihood of false positives.

Scope of availability: it is intended that all materials that inform how to use this tool including supporting data (datasets used for model development), training tools and other materials will be made freely available and housed in an appropriate repository.

Background information as submitted by the applicant[i] (summarised)

A letter of intent was sent to the EMA on February 10th, 2010, requesting scientific advice. Following telephone and in person advice with EMA, a briefing package was subsequently forwarded to the EMA containing specific questions for advice and an updated analysis plan, and a meeting was scheduled for September 1st, 2010. A written response and clarification questions were issued by the EMA to CAMD on August 23rd, 2010. CAMD provided written responses to the EMA on August 30th, 2010.

The scientific advice meeting was held with the SAWP on September 1st, 2010. The qualification advice in written by the CHMP was on January 19Th, 2011. The discussion focused on the specific questions that had been supplied by CAMD and those clarifying questions that had been supplied by the EMA. Specific concerns raised by the EMA were primarily ensuring that sufficient types of data informed the model, and that appropriate external validation be completed to ensure that the model adequately described the data. EMA also explained the need for the subsequent submission to take the form of specific issues and questions. In general, the EMA was very supportive of the approach.

A complete submission dossier was submitted to EMA on March 20th, 2013. This was followed by a teleconference between the CAMD team and the SAWP on May 7th, 2013. During this meeting, the proposed context of use was presented; the major components of the model were discussed (disease progression, drug effects, placebo effect and drop out functions) along with the relevant covariates for each of these; the internal and external validation results were shared; and an example on the application of the model was explained. Final meeting minutes were sent to EMA on May 10, 2013 (CAMD meeting minutes, dated May 10, 2013). EMA additional questions were received on May 17, 2013.

A face-to-face meeting was held with SAWP on June 4th, 2013. The discussion focused on answering the twelve questions that EMA sent following the May 7th, 2013 teleconference. Final meeting minutes were issued on June 10th, 2013 (Section 4.9. CAMD meeting minutes, dated June 10, 2013).

Specific questions for EMA review

CAMD believes that adequate data and analyses are presented to justify a regulatory decision on qualification of this tool. CAMD believes the questions below are the core questions that need to be addressed in order to ascertain the suitability of the model for qualification. CAMD requests responses on the specific questions pertaining to the submitted longitudinal model describing changes in cognition

in patients with mild and moderate AD. The purpose of the model is for use in clinical trial simulations of various designs using ADAS-Cog in the mild and moderate AD population, to allow for trial optimization, and to provide a quantitatively-informed background from which sponsors can work when designing implementing, and evaluating trials for individual compounds.

1) DATA. Does the Agency agree that

a) The endpoint selected (ADAS-Cog) is suitable for describing cognitive changes in mild and moderate AD?

b) The data being used (literature, ADNI, and CAMD database) are sufficient to describe longitudinal changes in ADAS-Cog in patients with mild and moderate AD?

2) MODEL. Dose the agency agree that

a) The proposed model provides an adequate quantitative longitudinal description of the progression of cognitive changes in mild and moderate AD for data from various sources? Specifically,

i) Changes in disease progression based on baseline severity have been adequately described?

ii) Changes in disease progression due to other patient factors (ApoE4 status, gender, age) have been adequately assessed in model development?

iii) Time dependent changes in variance have been adequately described?

iv) The placebo effect described by the model is consistent with current clinical opinions?

v) Symptomatic agent effects described by the model for acetylcholinesterase (AChE) inhibitors are consistent with current clinical opinions?

vi) The predictive checks and external validation are sufficient for use for trial simulation purposes?

3) SIMULATION. Does the Agency agree that

a) A simulation approach based on a quantitative model is an adequate strategy for the purpose of comparing clinical trial designs with cognition as a primary endpoint in mild and moderate AD?

b) The example simulations provided in the submission are sufficient to demonstrate the utility and use of this model as a drug disease trial (DDT)?

4) Does the Agency agree that this DDT, as defined in this document, is suitable for qualification for use in drug development as a longitudinal model for describing changes in cognition in patients with mild and moderate AD, and for use in assisting in trial design in mild and moderate AD, as defined by the context of use?

Background and rationale for use of an ADAS-Cog model in AD as a disease progression and trial evaluation model

ADAS-Cog scale

The Alzheimer's Disease Assessment Scale was designed to measure the severity of the most important symptoms of Alzheimer's disease (AD) (Mohs et al., 1983). Its subscale, ADAS-Cog, is the *de facto* standard primary outcome neuropsychological measure for AD trials (Rosen et al., 1984). It consists of 11 tasks measuring the disturbances of memory, language, praxis, attention and other cognitive abilities which 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. Elderly normal adults may score as low as 0 or 1, but it is not unusual for nondemented adults to score slightly higher. It has been, and is routinely used in global clinical trials of patients with mild and moderate AD.

The ADAS-Cog has been the primary cognitive endpoint used for US approvals for all past and currently marketed compounds labeled for the treatment of mild and moderate AD including tacrine, rivastigmine, galantamine, and donepezil. (Note that clinical trials that evaluated memantine, which is indicated for moderate to severe AD, utilized the severe impairment battery [SIB]).

To our knowledge, the ADAS-Cog is the agreed primary cognitive endpoint for all recent global phase II and phase III drug development programs in patients with mild and moderate AD.

Disease progression and trial evaluation models

In designing a clinical trial, the interested parties gather information from a multitude of sources. They may use past and recent literature to inform them about expected treatment effects and current study designs in use. They may have patient-level data in their organization that informs them about expected intra subject, inter subject variability, and inter occasion variability. They often have past clinical trial experiences that they draw from (which varies between individuals). The team designing a clinical trial will attempt to implicitly integrate all of this information to form conclusions about what design is likely to be best for the stage of development and the compound in question.

A model based-trial simulation does exactly the same thing. The only difference is that the data being used, the model and its assumptions, as well as the scenarios to be tested to determine optimal design are all explicitly defined.

Alzheimer's disease modeling for use as a disease progression and trial evaluation model

Assumptions about disease progression and the time-variant effects of placebo and existing drug treatments for AD form the basis for various decisions made in AD drug development, including decisions relating to trial design and analysis. While ad hoc synthesis of estimates from a small number of trials can, in some cases, form sufficient evidence base for such assumptions, it is generally a more informative and objective approach to concisely summarize all available and relevant data with the aid of a meta-analytic model. Such a meta-analytic synthesis is particularly relevant in Alzheimer's disease, where extensive historical data are available. Moreover, models may be used to interpolate expected results and to simulate data under conditions that have not been previously studied, e.g. when sampling at different time points or when enrolling patients with a different set of covariates.

Clinical trial simulations in AD

Clinical trial simulation is a means of estimating relevant operating characteristics for essentially any clinical trial design under any hypothesized parameter configuration for the "true" effects of a drug. It may be used to assess how different trial design and drug factors affect trial performance. These factors may be controllable trial design properties, such as the doses studied, the sampling times, the optimal study duration and sampling times, and use of washouts (Hennig et al., 2009) or uncontrollable factors, such as the drug characteristics (pharmacokinetic or pharmacodynamics) (Lockwood et al., 2006). Other influencing factors may include the progression of disease over time or subject specific characteristics that may be related to disease progression or treatment response.

In some cases, simulation may not be needed to assess certain operating characteristics (such as power) where conventional analytic approximations are available. However, these analytical approximations are not available for all trial designs of interest or all model assumptions of interest, and even in these cases, a simulation-based assessment may be more accurate (assuming a sufficient number of simulations are used). Provided there are specific decision rules for determining that a

particular trial was positive, or for judging an estimate to be sufficiently accurate, clinical trial simulation can also provide a rational basis for making decisions about the trial design and quantitating how effectively the design can answer the study objectives (Bonate, 2000; Holford et al., 2000). Clinical trial simulation can be viewed as an extension of conventional statistical design evaluation. Data derived models are utilized based on the relationship between dose, exposure, the time course of disease progression, placebo effect, and the outcome measure, providing an alternative approach to that described in the statistics literature (Putt and Ravina, 2002).

Summary of disease progression models for ADAS-Cog to date

Historical AD models

Various disease progression models for clinical outcomes in AD have been published (Holford and Peace, 1992; Chan and Holford, 2001) and the methods have been well described (Mould et al., 2007). Past work has focused on ADAS-Cog, which is the primary endpoint for cognition in nearly all clinical trials in mild and moderate AD. While the general model building principles and model structure provided similar results and interpretations, the studies upon which these models were based were of short duration, as little as 12 weeks, and did not contain newer key data such genotype information, now shown to be an important covariate in understanding the rate of progression of AD and the rate of cognitive decline in AD patients (Atchison et al., 2007). In addition, these models lacked certain structural features that would improve their use for clinical trial simulation, such as constraining the limits of the ADAS-Cog (zero to seventy), and allowing for variance components to change over time (an essential feature if the model is to be used for clinical trial simulation of disease progression for AD).

Model based AD meta-analyses

More recently, Ito et al (Ito et al., 2010) applied a model-based meta-analysis to summary level data available in the literature, to quantify the dependence of rates of progression on baseline ADAS-Cog scores. In this analysis, a systematic literature review from 1990 to 2008 for all available AChE inhibitor studies, as well as clinical studies that evaluated the rate of deterioration in AD patients was conducted. From 52 trials, which represent approximately 19,992 patients and more than 84,000 individual observations, a total of 576 mean ADAScog change–from-baseline data points were collected. Based on the data available from these articles, a model was developed to describe the longitudinal response in ADAS-Cog (change from baseline) in mild-to-moderate severity AD patients. The model described the rate of disease progression, the placebo effect observed, and the symptomatic effect of AChE inhibitors. Baseline ADAS-Cog, mini-mental state examination (MMSE), age and publication year were tested as covariates.

Ito's model reports that disease progression in mild-to-moderate AD patients across all available and relevant literature sources was estimated at 5.5 ADAS-Cog units per year. An Emax-type model best described the symptomatic drug effect for AChE inhibitors. The rate of disease progression (underlying disease progression) was not different between placebo and AChE inhibitor treated groups. Ito's model identified baseline ADAS-Cog as significant covariate on disease progression. Baseline age was also tested as a covariate on the rate of disease progression but the model was not able to describe any effect, likely due to the narrow distribution of mean age (literature-level analysis). There was no significant impact of publication year in the model.

The literature based meta-analyses provided a useful and complete integration of the estimated natural history of AD and provided estimates of treatment effects for currently available AChE inhibitor therapies. However, due to the nature of the literature data in that it is only study-level summary data; the model had limited ability to evaluate important individual covariates, such as age and ApoE4 genotype. Also, the meta-analysis model from the literature using study-level data neither provides inter-subject variability information nor includes components for variance increasing over time.

Ito ADNI model (2010)

In 2010, Ito et al published a patient-level model-based meta-analysis to describe the longitudinal response in ADAS-Cog obtained from the Alzheimer's disease neuroimaging initiative (ADNI) (Ito et al., 2011). The model was fit to the longitudinal ADAS-Cog scores from 889 patients. Risk factors (age, ApoE4 genotype, sex, family history of AD, and years of education) and baseline severity were tested as covariates. Results indicated that rate of disease progression increased with baseline severity. Age, ApoE4 genotype, and sex were identified as potential covariates influencing disease progression. The rate of disease progression as described by the ADAS-Cog in mild-to-moderate AD patients was estimated at approximately 5.5 ADAS-Cog units/year, similar to that reported using literature based analyses.

The authors concluded that a linear disease progression model adequately described the natural decline of ADAS-Cog observed in ADNI over 2-3 years within the individual patients. Baseline severity, which is incorporated into the model to explain the non-linearity of the disease progression, is an important covariate to predict a curvilinear rate of disease progression in normal elderly, mild cognitive impairment (MCI) and patients with Alzheimer's dementia. Age, ApoE4 genotype, and sex also influenced the rate of disease progression.

Faltaos model

In April 2011, Faltaos et al presented "Quantification of disease progression and drop-out for Alzheimer's disease" at the American conference on pharmacometrics (ACoP) as a poster and podium presentation in San Diego, CA (William-Faltaos et al., 2013). This work was supported through a fellowship within FDA, and funded by the American association of pharmaceutical scientists (AAPS). The research project aimed to quantitatively describe the natural progression of Alzheimer's disease (AD) based on the ADAS-Cog in patients with mild-to-moderate AD using prior trial data. Data from 10 placebo-controlled clinical trials including more than 2400 patients with mild and moderate AD with up to 72 weeks of treatment were used. Different models describing the time course of ADAS-Cog score were evaluated. Patient characteristics (age, gender, race) that could potentially affect the score changes were assessed, but none were identified (see below). In addition, patient drop-out patterns were characterized using parametric survival models. Covariate selection was further performed to identify the risk factors associated with a higher drop-out rate.

The time course of the ADAS-Cog in patients with mild and moderate AD receiving placebo was best described by a log linear model, where the intercept represents the log-transformed ADAS-Cog score at week 10, and the slope is the disease progression (i.e., natural increase of ADAS score) on the log scale.

Covariates influencing the intercept were baseline ADAS-Cog score and baseline MMSE. No covariates influenced the disease progression slope. A parametric log-normal model fitted the dropout data best. Baseline ADAS-Cog and age were found to be significant predictors for dropout.

Samtani ADNI model

In April 2011, Samtani et al presented "An improved model for disease progression in subjects from Alzheimer's disease neuroimaging initiative" at the American conference on pharmacometrics (ACoP) meeting in San Diego, CA. The complete work is now available as a publication from the Journal of clinical pharmacology (Samtani et al., 2012). The objective of this analysis was to develop a semi-mechanistic non-linear disease progression model using an expanded set of covariates that captures the longitudinal change of ADAS-Cog scores from the ADNI study that consisted of 191 patients with mild AD who were followed for two years. The model described the rate of progression and baseline disease severity as a function of influential covariates. The covariates that were tested fell into 4 categories: a) imaging volumetric measures; b) serum biomarkers; c) demographic and genetic factors; d) baseline cognitive tests.

Covariates found to affect baseline disease status were years since disease onset, hippocampal volume and ventricular volume. Disease progression rate in the model was influenced by age, total serum cholesterol, ApoE4 genotype, trail making test (part B) score, as well as current levels of cognitive impairment as measured by ADAS-Cog. Rate of progression was slower for patients with mild and severe AD compared with moderate AD.

Conclusions from previous model based experiences

The features of the models described above are compared in table 2. In general, all models developed to date to describe ADAS-Cog have utilized similar basic concepts of disease progression. That is, describing the ADAS-Cog at any given time as a function of the baseline score, and a progression of the disease as a function of time. Historical models described changes in AD as a linear progression. The Ito literature model identified that the severity of the disease itself influenced the slope, and thus the slope changed over time (introducing non-linearity). More recent models directly use non-linear relationships to describe the course of disease over time.

The models described here have utilized a variety of data types including summary level data from literature sources, data directly from one or more of a related series of controlled clinical trials, or non-interventional natural history studies.

The models described in the literature also vary with respect to the component that would be required for a drug-disease-trial model. Of those described in table 2, none currently have all three components as described below:

Disease-drug-trial models can be described as follows (Gobburu and Lesko, 2009):

1. A disease model: Such models quantify the natural longitudinal progression of the outcome, the relationship of biomarkers to the clinical outcome, and the placebo effect observed within trials.

2. A trial model: Such models incorporate components of what is known about the patient population (baseline disease severity, etc.), patient drop-out rate and factors impacting it, as well as patient therapeutic adherence.

A drug model: Such models incorporate what is known about a compound's effectiveness, the impact of patient characteristics on drug effect, and changes in drug effect(s) over time.

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 to inform underlying disease progression, placebo arm data to inform about magnitude, onset and offset of placebo response in controlled clinical trials, estimates of various drug effects (magnitude, time to onset, and durability), rate and magnitude of drop-outs in the trials, and a rich source of covariates for model building. Unfortunately, no single trial can provide all of these elements.

Table 2. Comparison of recent ADAS-Cog longitudinal models in the literature

Model	Drug Effect Component	Trial Component	Data Source	Covariates	Linearity
Historical	Yes	Varied	Individual studies (tacrine)	Varied	Linear
Ito Literature	Yes (symptomatic agents estimated)	Placebo (onset and magnitude)	All controlled studies in the literature 1990-2008	Baseline severity	Linear (non-linearity introduced by baseline

					covariates)
Ito ADNI	No (NA)	No (NA)	ADNI (normal, MCI, mild AD)	Baseline severity Age, ApoE4 genotype, and sex	Linear (non-linearity introduced by baseline covarites) Fits normal MCI and mild AD
Samtani ADNI	No (NA)	No (NA)	ADNI Mild AD	Disease onset, hippocampal volume, and ventricular volume, age, total cholesterol, APOe4 genotype, trail making test (part B) score	Nonlinear Fits mild AD
Faltaos	No	Drop-out No Placebo		Covariate influencing the intercept were baseline ADAS-Cog score (did not use data prior to 4 months) and baseline MMSE score. No covariates influences the disease progression slope	Nonlinear (log transform not suitable for whole range of ADAS-Cog scores of 0-70).

Methodology

Given the success of previously published models in characterizing many aspects of the progression of ADAS-Cog values, CAMD's intent was utilize key learnings from existing models and adapt them in a manner that would support a comprehensive meta-analysis and that would enable realistic clinical trial simulation. The present effort focused on issues of estimation, demonstration of model validity, and examples of application. 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 "adopted" features included:

1. The use of a generalized logistic function to describe the natural progression of the disease on a constrained scale (Gillespie, 2009).
2. The use of a Bateman-type function to describe the incremental placebo effect (Holford and Peace, 1992, Ito et al., 2010).

3. The use of Emax functions to describe the incremental effects of approved AChE inhibitors as a function of dose (Ito et al., 2010, Gillespie, 2009).

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 (Ito et al., 2010, Samtani et al., 2012).

5. The use of baseline age and baseline severity as covariates on the hazard of drop-out (William-Faltaos et al., 2013).

In addition, CAMD has incorporated a number of important innovations:

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 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. The use of the logistic function is itself only sufficient to ensure that conditional expectations respect the 0-70 constraints. However, when the generalized logistic function is used with normally distributed residuals, there is a positive probability of simulating results outside of the allowable range. The Beta-distribution eliminates this.

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).

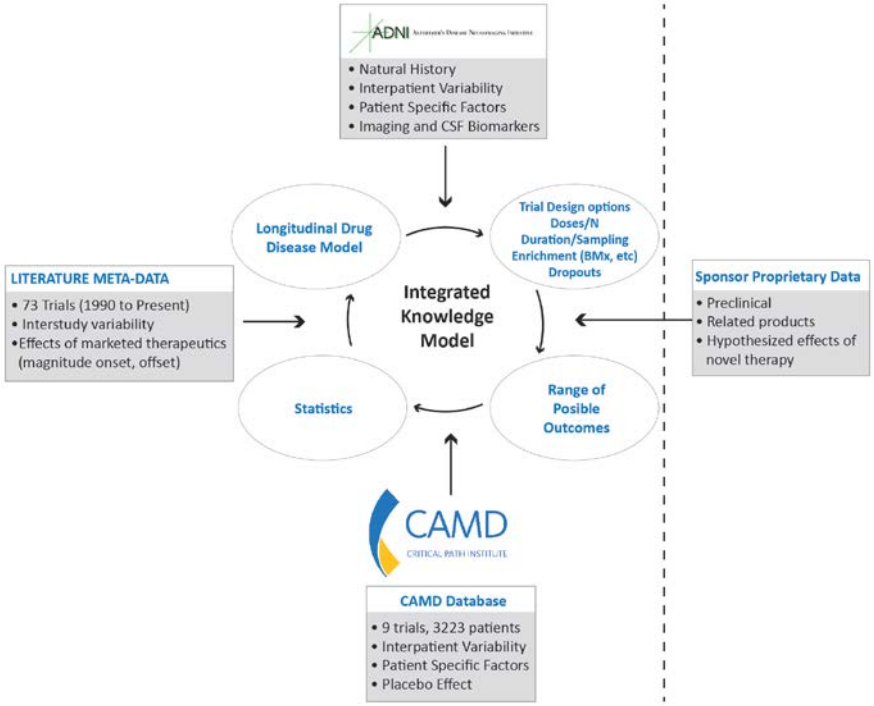
Modifications to planned analyses

The analysis plan submitted to FDA and EMA as part of the briefing package had included an extensive matrix of simulations to compare trial designs for use in AD, and to provide information for various types of expected treatment effects.

Data collection

In this analysis, data from three sources was utilized to inform model development (figure 2). ADNI data provided a rich source for the natural history of disease progression in patients with mild AD, and the most complete source of imaging and biomarker data collected to date in any AD trial. The CAMD database provided a rich source for individual level control arm data in mild to moderate AD patients (both placebo and background therapy). The literature (which provides summary level data) provided data for the model to estimate symptomatic treatment effects for AChE inhibitors, long term disease progression in controlled mild to moderate AD trials, and inter study variability.

Figure 2. Illustration of data sources used for model development in the submission



Literature data and selection criteria

A full description of the literature selection criteria is included in the unabridged briefing package, submitted March 21, 2013.

In summary, literature was searched and selected according to the approach suggested at the quality of reporting of meta-analysis (QUOROM) conference. A systematic search of public data sources (Medline, Embase, NICE and Summary for Basis of Approvals at FDA) from January 1990 to December 2010 was conducted. Key search terms were: AChE inhibitor generic names (donepezil, galantamine, rivastigmine, tacrine, velnacrine), trial endpoints (ADAS-Cog, MMSE, CIBIC, etc.), and clinical trial design descriptions (double-blind, randomized, controlled etc.).

ADNI data

All ADNI data used in this submission were obtained from the Alzheimer's disease neuroimaging initiative (ADNI) database (www.loni.ucla.edu\ADNI).

The data used in these analyses are derived from the data available from ADNI June 1st, 2010 (www.loni.ucla.edu\ADNI).

CAMD database

Consensus was reached on how best to share patient-level control arm data from CAMD member companies, in order to develop an AD precompetitive data repository. It was agreed the repository would align with the CDISC study data tabulation model (SDTM) industry data standard, since pharmaceutical companies will align with this in submissions to FDA, as the foundation for standardized clinical content. The focus for new standards was on the ADAS-Cog and MMSE in CDISC SDTM.

The data standards and integration workgroup of CAMD worked with the modeling and simulation work group to better understand the needs for standard data elements to fulfill the model development effort. 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.

Through periodic interactions, the group aligned the scales utilizing CDISC spreadsheet templates, in order to define the database table structure with the associated terminology. Each scale domain was reviewed by the CDISC submission data standards team.

The next step involved the CAMD sponsors mapping their respective studies from their source database structure to the CDISC SDTM data domain. The mapping process involved each company progressing through a learning curve on the standards. Mapping of the legacy data to the new standards involved programming to restructure the source data to meet the SDTM domain structures and also include the SDTM approved terminology for data values.

The effort took on average two months per sponsor to complete.

CAMD chose Ephibian, an organization based in Tucson, AZ, as the database and user interface developer based on demonstrated experience. Open-source SDTM-based validation software was integrated in the system to automatically validate incoming data. Each validation report was reviewed for SDTM compliance and fitness for the database. Datasets were either approved to the production database, or sent back for corrections to the supplier. The group is currently in the process of transforming data from clinical studies from academic sources for future inclusion in the CAMD database as well.

Table 5 describes the 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 in the model-development and evaluation process, and thus are not included in table 5. The studies included in the CAMD database consist of all control arm data from all 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 are not available in the database.

Disease progression and drug effect model development (abridged)

A full description of the model development is available in the unabridged edition submitted March 21, 2013. A technical description of the model is also available as a peer reviewed journal article (Rogers et al., 2012b).

Clinical trial simulations

Several clinical trial simulations were run for illustrative purposes. These simulations are not intended to provide evidence toward any global preference of a particular design. On the contrary, the intent is to suggest how a development team might use the model and associated simulation tools to select designs that are tailored to particular assumptions about the magnitude, onset, and offset of drug effects. For this purpose, several hypothetical scenarios were envisioned:

1. **Symptomatic drug effect scenario.** The drug properties assumed in this scenario are qualitatively similar to those of marketed AChE inhibitors, albeit with some modifications to make the interpretation of the example more straightforward. The onset of effect is assumed to have an Emax functional form with a mean (placebo-adjusted) effect of 2.275 point change in ADAS-Cog at 24 weeks, an ET50 of 1 week, and a half-life for offset of effect (after discontinuation of treatment) of 1 week. The candidate designs considered in this example were:

    a. A parallel design (75 patients per arm) with 12 week treatment duration, assessments at weeks 0, 1, 3, 6, 9, and 12. The envisioned primary analysis is based on a linear mixed effects model with random subject effect and fixed effects for baseline ADAS-Cog, visit (nominal scale), treatment, and visit by treatment interaction, with the

treatment comparison formulated as the expected difference at week 12 (using interaction contrasts).

b. A cross-over design (30 patients per arm) with two 6 week treatment durations and a 3 week washout period in between. Assessments within each treatment period were envisioned at weeks 0, 1, 3, and 6. The assumed primary analysis is based on a linear mixed effects model with random subject effect and fixed effects for baseline MMSE stratum, period, sequence, treatment, relative week (within period, nominal), period by relative week, and treatment by relative week. Treatment comparison was formulated as the expected difference at (relative) week 6 (using interaction contrasts).

For compounds where pre-clinical data suggests that only a symptomatic effect is likely to be observed, the key objective in early studies in patients is to determine whether the proposed mechanism translates into meaningful changes in a clinical outcome, as rapidly as possible. Often in early development, duration of toxicology exposure, and concerns for patient safety push teams to explore short, rapid proof of concept (POC) designs. Therefore, exploring the optimal POC studies is of interest.

The example will provide an average simulated trial for 6 week cross-over design and 12 week parallel design, respectively under a symptomatic drug scenario. Under these assumptions used for this simulation, treatment effect (difference between placebo and treatment) at the end of each 6 week period was independent of treatment period in the cross-over design. In this context, a cross-over design has the potential to reduce the sample size while maintaining appropriate power to demonstrate the drug benefit.

In order to have a fixed point of comparison in the evaluation of both designs, the “true effect” of the drug was formulated as the 2.275 point difference at 24 weeks. This corresponds to a common scenario in early phases of drug development in which the estimand of interest is an effect at a time point later than any of the time points studied. From this perspective, some bias is expected with both designs, since the full drug effect is not attained over the duration of the study.

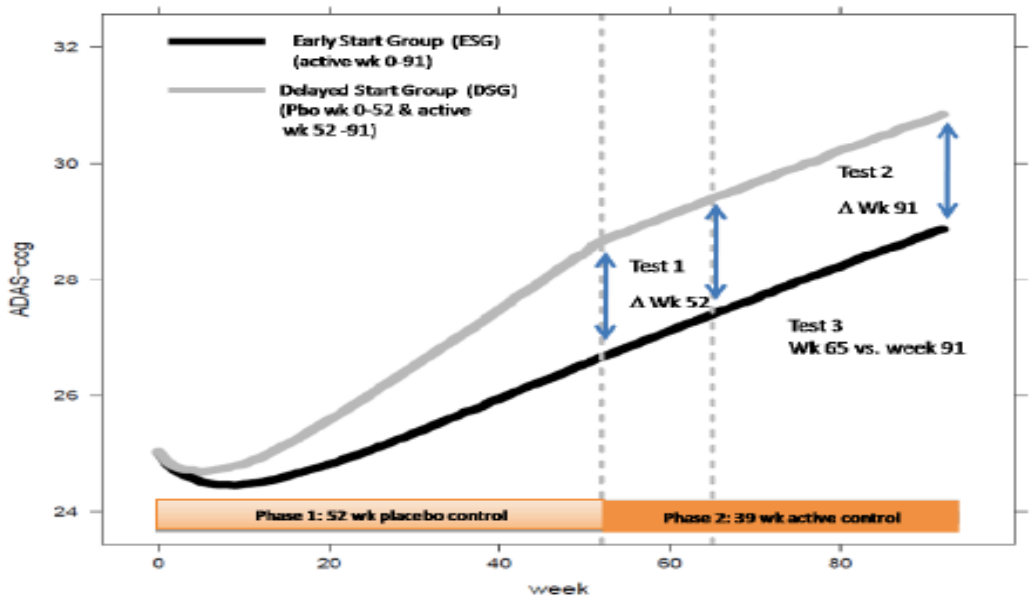
2. **Disease modifying drug effect scenarios.** Hypothetical drug effects were expressed as proportional reductions in expected progression rates. Based on the feasibility to detect a potential effect, the proportional reductions considered were 5%, 10%, 20%, 30%, 40%, and 50%. The candidate designs considered in these scenarios were:

    a. A parallel design with 78 week treatment duration and assessments at weeks 0, 26, 52, and 78. The assumed primary analysis used a multivariate model for repeated measures (MMRM) approach with unstructured covariance matrix and fixed effects for baseline ADAS-Cog, treatment, visit (nominal), and treatment by visit interaction. Treatment comparison was formulated as the expected difference at week 78 (using interaction contrasts).

    b. A delayed-start design (D’Agostino 2009, Olanow et al., 2009). This design employs a placebo-control phase (phase 1), and an active control phase (phase 2). The patients who receive placebo in the placebo control phase and study drug in the active control phase are referred to as the delayed-start group. The patients who receive study drug in both phases are referred to as the early-start group. Fifty-two week and 39 week duration was selected for phase 1 and phase 2, respectively with the final 26 weeks being used to assess stability of effect (in the notation of (D’Agostino 2009) T2 = 52, T3 = 65, T4 = 91, and T1 is not relevant for our purposes because our envisioned

primary analysis does not invoke slopes or assume linearity with respect to time). Assessments were assumed at weeks 0, 26, 52, 65, 78, and 91. A schematic for this design is provided in figure 3.

Figure 3. Schematic of delayed start design for a disease modifying agent



The envisioned primary analysis would test the three research hypotheses associated with delayed start designs.

1. Test for difference in ADAS-Cog change from baseline between the placebo and study drug group at end of phase 1 (52 week).
2. Test for difference in ADAS-Cog change from baseline between early and delayed-start groups at end of phase 2 (91 week).
3. Test for evidence for the stability of the treatment difference, which may be assessed by comparing the change from week 65 to week 91 for early versus delayed start groups.

The formulation of any of these three hypotheses in terms of slopes is confounded given that the present model implies non-linearity of the time courses. Consequently, CAMD envisions the sponsor testing all three hypotheses using interaction contrasts rather than slopes, using the same MMRM model as described for the 78 week parallel design described before. Since there is no consensus regarding an appropriate equivalence margin for testing the stability of effect (whether formulated as a slope or an interaction contrast), the typical values for a 90% confidence interval are reported that could be used in an equivalence test.

For each design 10,000 trials were simulated in order to estimate operating characteristics. Each simulation iteration proceeded as follows:

1. Baseline MMSE entry criteria were specified and baseline MMSE values were generated uniformly within this range.
2. Baseline age values, ApoE4 genotype values, and gender values were simulated from the posterior (according the joint distribution for these covariates implied by the model).
3. Block randomization was used to assign each simulated patient to either treatment or placebo.

4. Longitudinal data was simulated for each patient using the model posterior in conjunction with the simulated covariate values and the treatment assignment for each patient. Each simulated study involved a separate draw form the distribution of random study effects. Drop-out times were simulated for each patient according to the dropout model posterior. Response values for visits occurring after the time of dropout were set to missing (for the majority of patients whose dropout time exceeded the trial duration, no values were set to missing).

5. A significance level of 0.05 (two tailed) was used to test the hypotheses. For each single simulated trial, a binary indicator of technical success (rejection of at least one null hypothesis) was captured in the simulation output. The proportion of simulated trials achieving technical success was then taken as the estimate of the statistical power of the trial design.

Results

Demographics

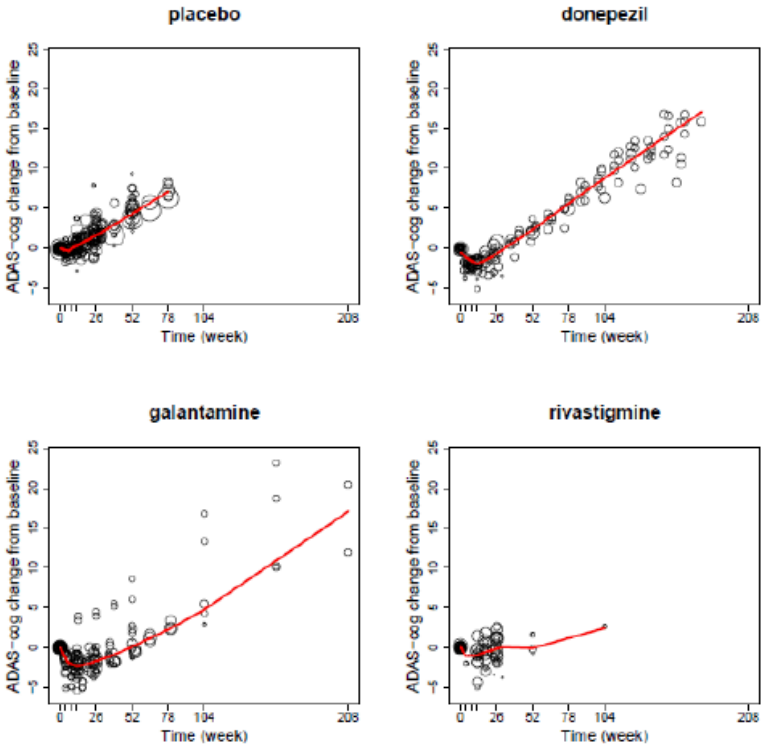
Literature data

Data from 73 studies were collected from the literature (Ref 36-106, derived from abridged version dated March 2013), 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). A brief summary of the characteristics of the trials collected in the literature database are also available in the unabridged version of this document. A full pdf version of each original article is also available on request.

Changes in the control arm data demonstrate a "*hockey stick*" shape, typical in most AD trials (figure 4 and figure 5). 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 offset the normal control arm data, but then return to the normal progression, maintaining an offset.

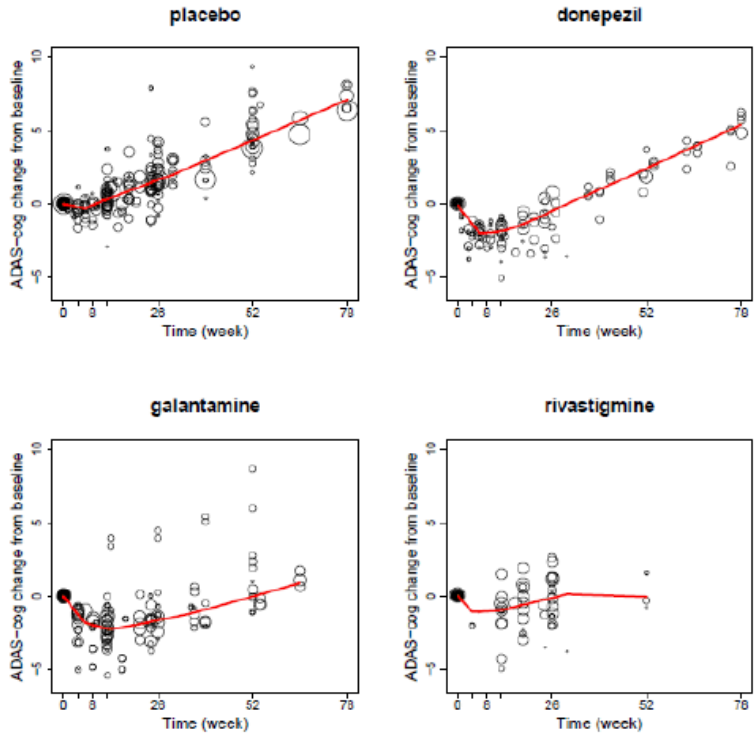
The relationship between baseline MMSE and ADAS-Cog obtained from the literature (figure 6, upper left panel) form CAMD studies (figure 11) and from ADNI (figure 14) appear similar.

Figure 4. Observed mean ADAS-Cog change from baseline over time by compound



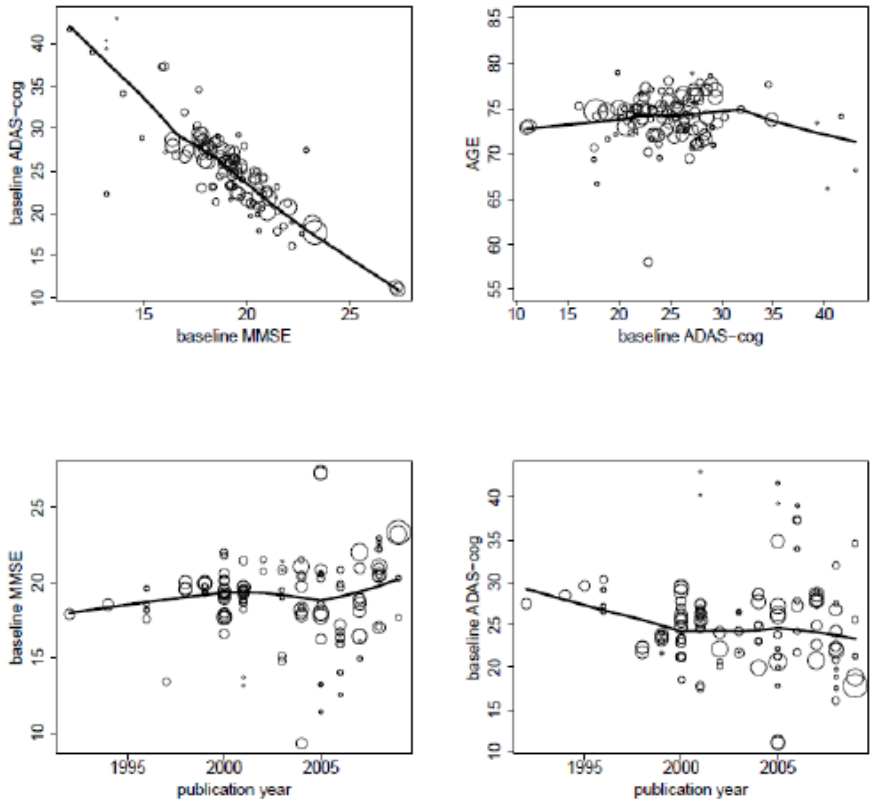
Source: ePharm artifact ID number 4925437
Loess line is in red.

Figure 5. Observed mean ADAS-Cog change from baseline over time by compound over 78 weeks



Source: ePharm artifact ID number 4925801 (Loess line is in red)

Figure 6. Observed mean ADAS-Cog change from baseline vs. Literature characteristics



Source: ePharm artifact ID number 528415

Patient level data

CAMD database

Basic demographics data were similar across the studies in the CAMD database (table 5). 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 one or two copies of the e4 allele) was also similar for those studies where this information was available.

ADAS-Cog scores and change from baseline ADAS-Cog scores are plotted by study (figure 7 and figure 8 respectively). ADAS-Cog scores and change from baseline ADAS-Cog scores are also plotted by baseline severity (figure 9 and figure 10 respectively). As can be seen, there is an apparent increase in the rate of disease progression as severity increases, as evidenced by the smooth lines fit to the data in these plots.

The relationship between baseline MMSE and ADAS-Cog obtained from the literature (figure 6) from CAMD studies (figure 11) and from ADNI (figure 13) appears similar.

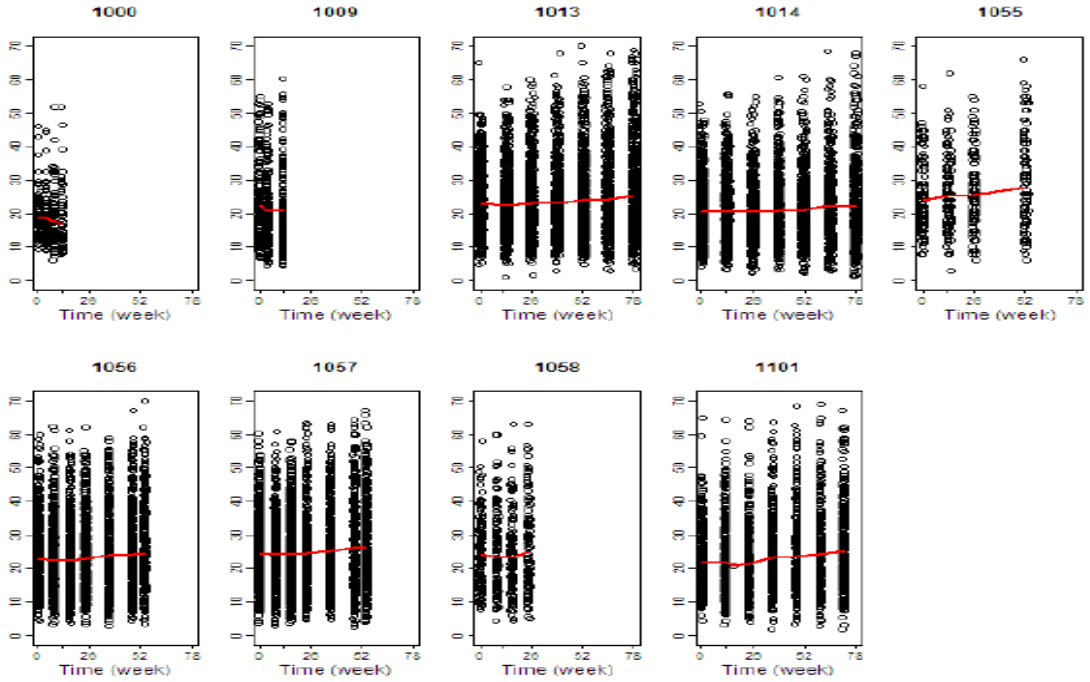
While the majority of patients in the CAMD database represent North America and Western Europe, all global major regions, including Asia, South Africa, and Latin America, are represented in the database (figure 12).

Table 5. 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 in yrs	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	61.0	59.3	51.1	**55.3**
APOE Status (%e4 Carriers*)	--	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)**
Yrs since diagnosis	2.6 (≤1-8)	<1 (≤1-11)	2.0 (≤1-10)	2.0 (≤1-11)	--	2.5 (≤1-20)	2 (≤1-10)	1.5 (≤1-10)	2.4 (≤1-12)	**2.0 (≤1-20)**
Stable background therapy (Yes/ No)	Yes	No	Both	Both	No	Yes	Yes	No	Yes	**Both**

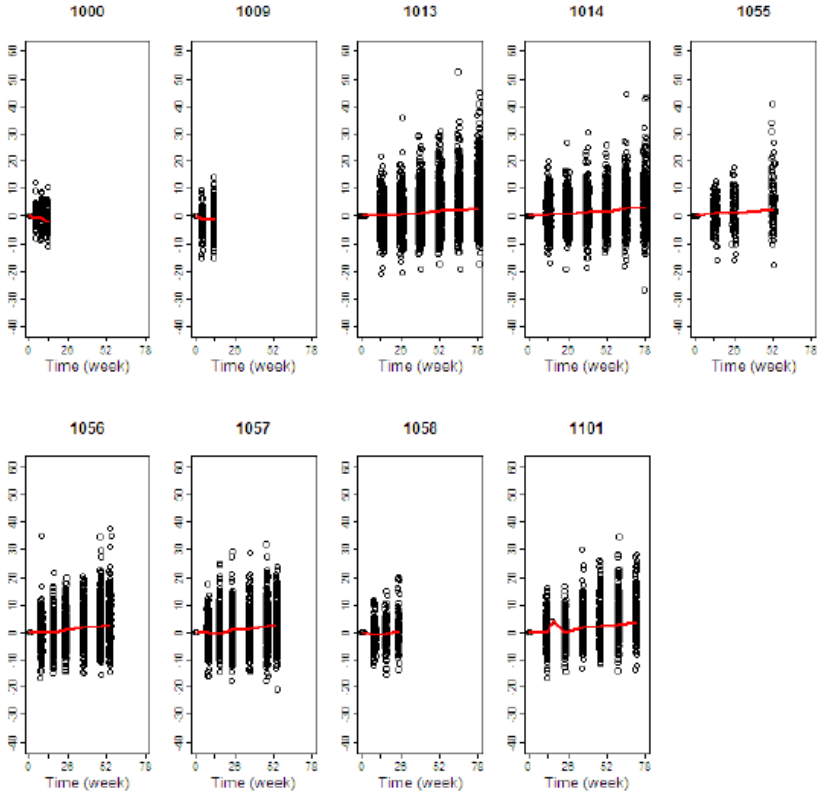
*APOE e4 carriers include patients with one or two copies of e4 allele at the APOE locus
(): standard deviation for age, bMMSE and bADAS-Cog, and range for year since diagnosis

Figure 7. Observed ADAS-Cog scores over time by study in CAMD studies



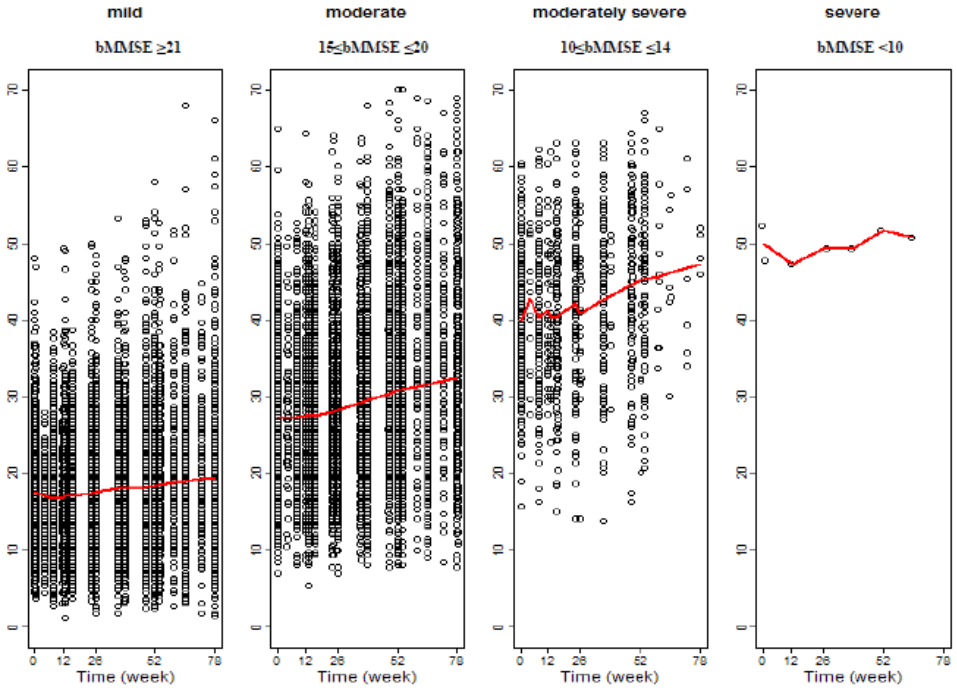
Loess line is in red.

Figure 8. Observed change from baseline ADAS-Cog scores over time by study in CAMD studies



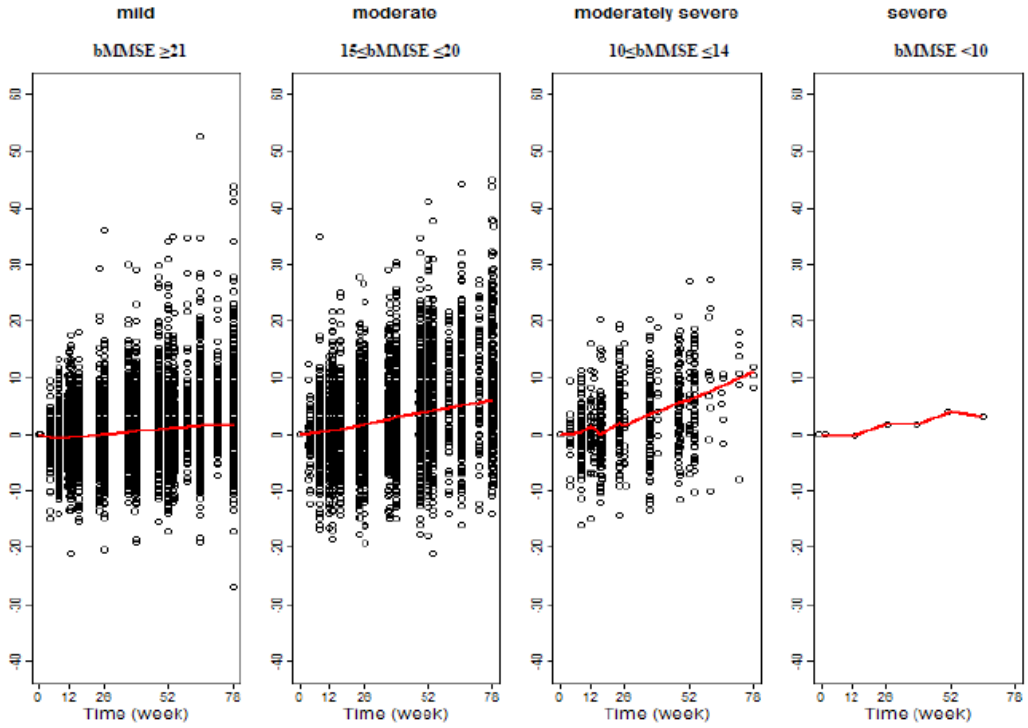
Loess line is in red.

Figure 9. Observed ADAS-Cog scores over time by baseline severity in CAMD studies



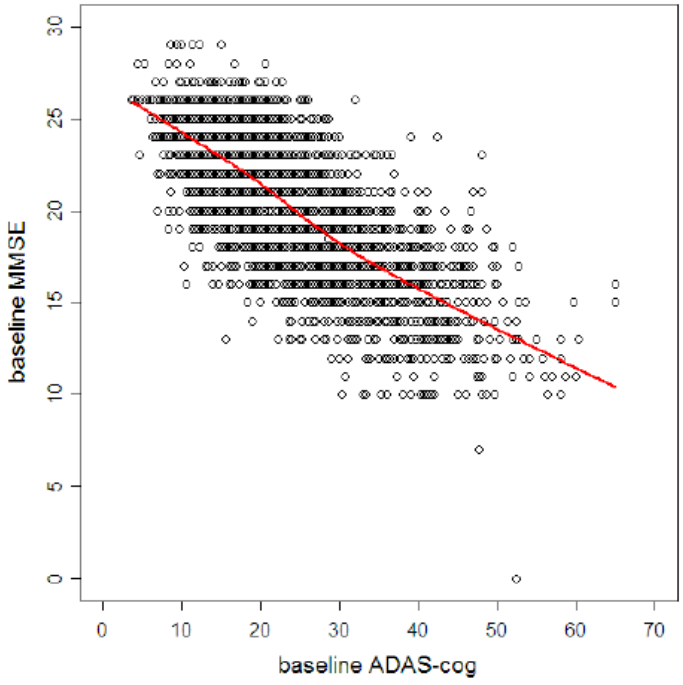
Loess line is in red.
*N=2 for severe patient group

Figure 10. Observed change from baseline ADAS-Cog scores over time by baseline severity in CAMD studies.



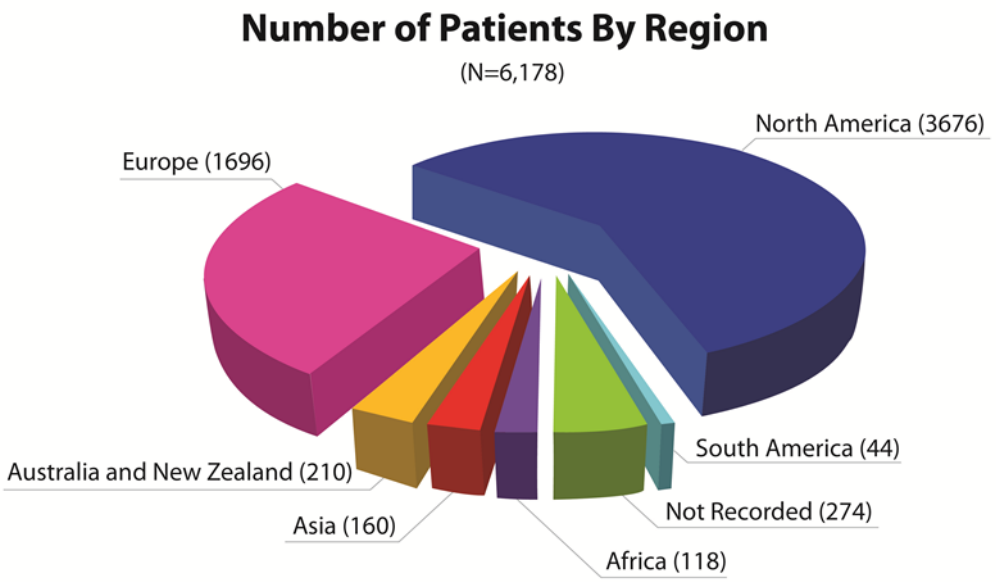
Loess line is in red.
*N=2 for severe patient group

Figure 11. Correlation of ADAS-Cog vs. MMSE in CAMD studies



Loess line is in red.

Figure 12. Number of patients by region in CAMD database



ADNI database

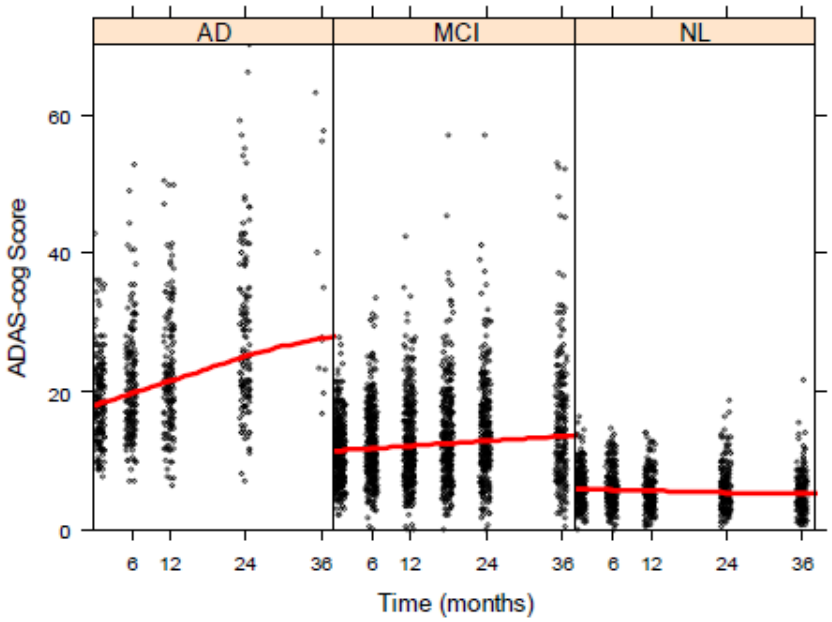
A complete description of the ADNI dataset available here is included, but only the AD patient data were used for the analysis. The dataset available contained 817 patients consisting of 229 normal (NL), 402 MCI and 186 AD patients (table 6). Overall, the age distributions are similar among these populations. The proportion of females in the MCI group is slightly lower but similar between AD and normal, with the majority of patients classified as white. The distribution of ApoE4 (Ɛ3Ɛ4 and Ɛ4Ɛ4) carrier status was more frequent in AD patients. Observed longitudinal ADAS-Cog data are visualized in figure 13 (line: loess) and the linear relationship between baseline ADAS-Cog and baseline MMSE is presented in Figure 14 (line: loess). As expected, baseline MMSE scores and baseline ADAS-Cog are highly correlated (figure 14). Because of the number of superimposed data points at the same time point, visit values (month) in figure 13 and actual score (MMSE) in figure 14 were slightly jittered in the figures to aid visual interpretation.

Table 6. ADNI demographic characteristics

	AD	MCI	NL
No of patients	**186***	**402**	**229**
Age (yr)	75.3 ± 7.6	74.8 ± 7.4	75.9 ± 5.0
Female (%)	47.3	35.6	48.0
Baseline ADAS-cog	18.7 ± 6.3	11.5 ± 4.4	6.2 ± 2.9
Baseline MMSE	23.3 ± 2.0	27.0 ± 1.8	29.1 ± 1.0
Education (yr)	14.7 ± 3.2	15.7 ± 3.0	16.0 ± 2.9
ApoE4 status			
ε4 non-carrier (%)	63 (33.9)	187 (46.5)	186 (73.4)
ε2, ε2 (%)	0	0	2 (0.9)
ε2, ε3 (%)	5 (2.7)	17 (4.2)	31 (13.5)
ε3, ε3 (%)	58 (31.2)	170 (42.3)	135 (59.0)
ε4 carrier (%)	123 (66.1)	215 (53.5)	61 (26.6)
ε2, ε4 (%)	4 (2.1)	11 (2.7)	3 (1.3)
ε3, ε4 (%)	83 (44.6)	157 (39.1)	53 (23.1)
ε4, ε4 (%)	36 (19.4)	47 (11.7)	5 (2.2)
Race (%)			
American Indian or Alaskan Native	0	1 (0.2)	0
Asian	2 (1.1)	9 (2.2)	3 (1.3)
Black or African American	8 (4.3)	15 (3.7)	16 (7.0)
White	174 (93.5)	376 (93.5)	210 (91.7)
More than one race	2 (1.1)	1 (0.2)	0

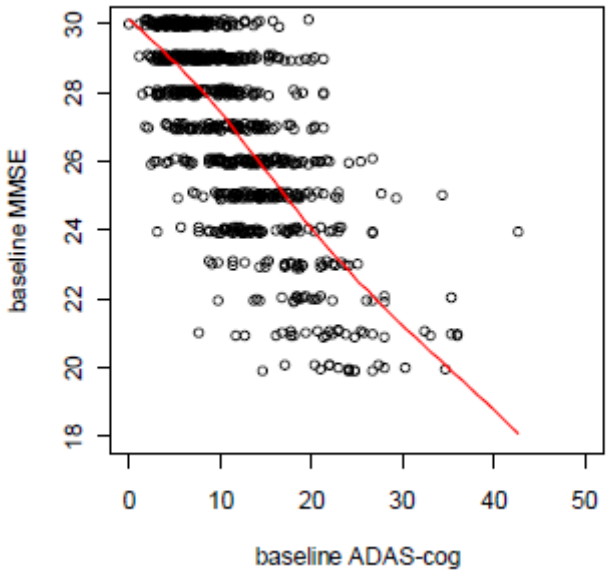
*; mild=171, moderate=13, severe=1, NA=1
Source: ePharm artifact ID number 280025

Figure 13. Longitudinal ADAS-Cog by patient population in ADNI study



Loess line is in red.

Figure 14. Correlation of baseline ADAS-Cog vs baseline MMSE in ADNI study



Loess line is in red.

Covariate model

In this document, the term “covariate model” is used to refer to the components of the model that describe the *distribution* of covariate values. The *effect* of covariates on the response is not considered to be part of the covariate model and is described instead as a component of the complete data model.

Convergence diagnostics

Convergence diagnostics for both covariate distribution parameters and complete data model parameters are provided in appendix (abridged in this version) 3.3.2.

Covariate model summary and evaluation

The final model included baseline MMSE, baseline age, ApoE4 genotype, and gender as covariates. While the effects of baseline MMSE are included in the model, the distribution of baseline MMSE was not itself modeled because:

Baseline MMSE was not missing from any records in the data set, so explicit modeling was not necessary for imputation purposes.

Clinical trial design teams generally exert a greater degree of control over the distribution of baseline MMSE values in a trial (e.g. via stratification) than they do over other covariates, so the notion of a “natural distribution” of baseline MMSE values within a trial is conceptually problematic.

Exploratory data analysis suggested that baseline MMSE was not correlated with age, gender, or ApoE4 genotype. Thus, from a simulation perspective, it appeared to be satisfactory to generate baseline MMSE values independently of the other covariates.

The joint distribution of baseline age, gender, and ApoE genotype is characterized in terms of both observed and model-predicted summaries of the joint distribution in table 7.

Table 7. ApoE4, gender, and age imputations

No. ApoE 4 Alleles	Relative Frequency		Proportion Male		Mean Age		SD Age	
	Observed	Predicted	Observed	Predicted	Observed	Predicted	Observed	Predicted
0	0.43	0.42	0.45	0.45	73.8	75.3	8.6	9.3
1	0.44	0.44	0.42	0.45	74.1	75.6	7.8	9.3
2	0.13	0.14	0.48	0.45	71.7	66.5	6.9	9.3

A potential dependence between age and ApoE4 genotype is plausible as a result of ascertainment bias: older homozygous carriers of ApoE4 may have been more likely to have advanced in the disease past the point where they could be enrolled consideration. Such a relationship is indeed suggested by both the observed and model predicted age distribution, although the predicted mean ages exhibited a greater dispersion both within and between genotypes than do the observed values. This discrepancy arises from the inferred covariate states for missing records. There were no missing records for gender in the individual level data, so for simplicity gender was considered independent of both ApoE4 and age for the covariate model.

Dropout model

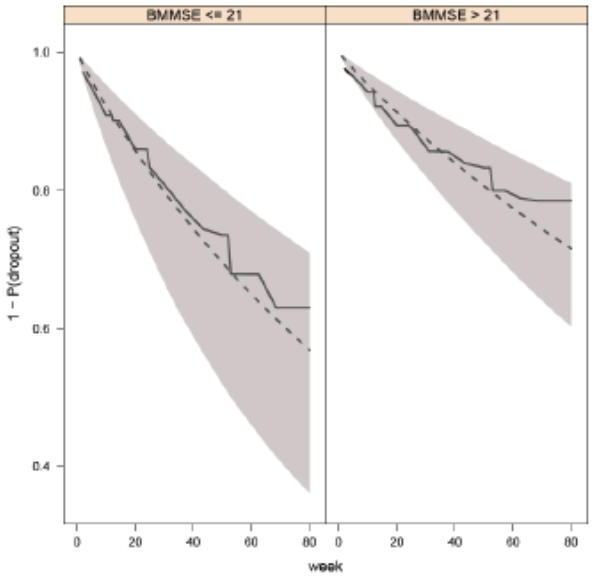
Convergence diagnostics for dropout model

Convergence diagnostic plots for dropout model parameters are provided in the appendix (abridged in this version) 3.3.2.

Dropout model summary and evaluation

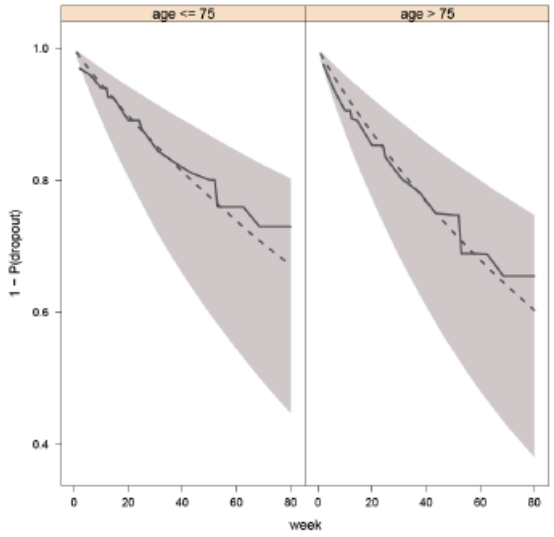
The fitted dropout model exhibited a high degree of agreement with the observed dropout rates, as seen in figure 15 and figure 16. The model predicted dropout rates as a function of time, baseline age, and baseline MMSE are summarized in table 8. The model adequately captures the dropout rate both by baseline MMSE and by age in these two plots.

Figure 15. Plot of probability (dropout) over time by baseline MMSE



Solid line represents Kaplan-Meier (non-parametric) estimates based on observed data; dashed line represents model prediction; grey region represents 90% credible interval for model prediction.

Figure 16. Plot of probability (dropout) over time by age



Solid line represents Kaplan-Meier (non-parametric) estimates based on observed data; dashed line represents model prediction; grey region represents 90% credible interval for model prediction.

Table 8. Model predicted dropout rates as a function of time, baseline age, and baseline MMSE

Covariate	Subset	Week	Median	5% LB	95% UB
Age	<= 75	26	0.133	0.0821	0.230
Age	<= 75	52	0.239	0.1530	0.393
Age	<= 75	78	0.329	0.2150	0.521
Age	> 75	26	0.165	0.1030	0.272
Age	> 75	52	0.293	0.1890	0.453
Age	> 75	78	0.398	0.2650	0.588
BMMSE	<= 21	26	0.179	0.1190	0.278
BMMSE	<= 21	52	0.315	0.2150	0.461
BMMSE	<= 21	78	0.426	0.2990	0.596
BMMSE	> 21	26	0.119	0.0809	0.166
BMMSE	> 21	52	0.216	0.1490	0.294
BMMSE	> 21	78	0.298	0.2110	0.401

Median indicates the posterior median (point estimate) for the drop-out rate, and LB and UB refer to the lower and upper bounds of the posterior credible interval.

Complete data model

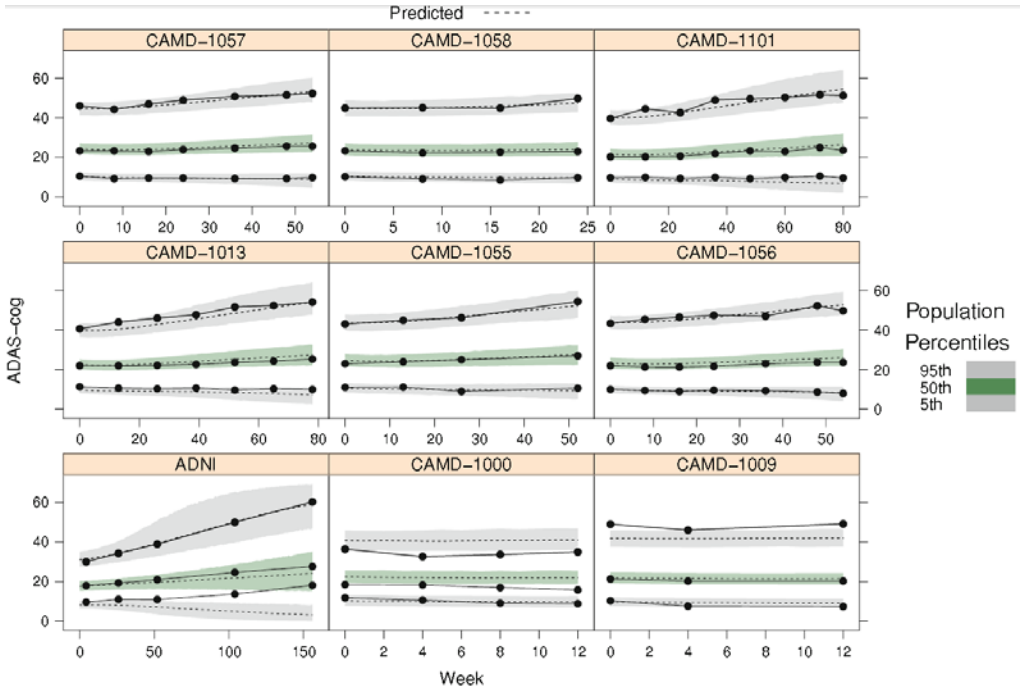
Convergence diagnostics

Convergence diagnostics for the complete data model are provided in appendix (abridged in this version) 3.3.2.

Posterior predictive checks (internal validation)

Figure 17 provides both the unconditional predictive checks by percentiles for studies that were included in the model building from the CAMD and ADNI datasets.

Figure 17. Unconditional predictive checks for sample population percentiles of ADNI and CAMD studies



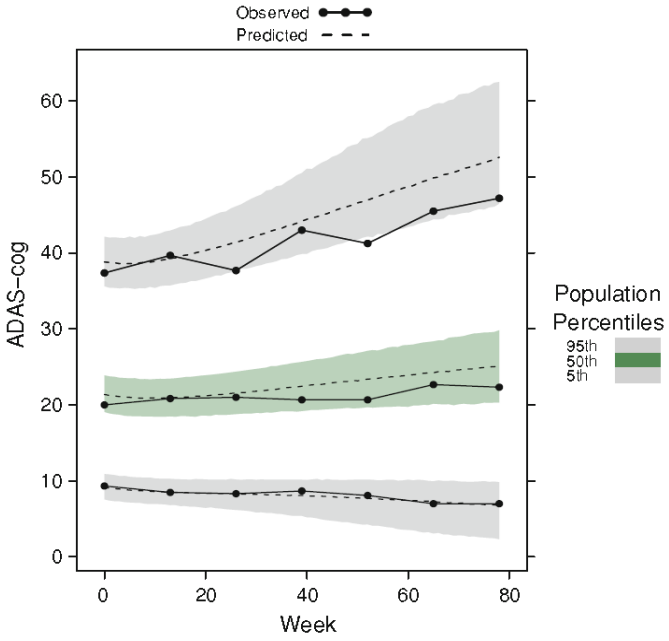
Black dots represent observed results, computed by nominal visit. Dotted line represents the posterior percentile model prediction and shaded region represents the 90% prediction interval when sampling from the posterior distribution with inter-study variation.

External validation

The external validation was conducted for the response to FDA request during the qualification review team meeting on April 28th, 2010. The response data from a randomly selected CAMD protocol (the *test set*) was withheld and blinded from model developers during the model development phase. The fitted model from the *model-building set* 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 (figure 18) to determine if all values fell within the 90% prediction interval.

Based on the criteria, the model met the external validation criteria that had been established. As such, the model correctly identified the results of this trial.

Figure 18 Unconditional predictive check for study 1014



Black dots represent observed results, computed by nominal visit. Dotted line represents the posterior percentile model prediction and shaded region represents the 90% prediction interval when sampling from the posterior distribution with inter-study variation.

Summary of model fit

The model developed provided parameter estimates similar to those described by previous authors. Estimates of baseline ADAS-Cog (intercept) from baseline MMSE (table 9) were consistent with the relationships observed between MMSE and ADAS-Cog in the literature (figure 6 upper left panel), CAMD (figure 11), and ADNI databases (figure 14).

The parameter estimates observed for covariates of age (table 10), baseline severity, and ApoE4 status on rate of disease progression on slope were also similar (table 11), yielding different rates of progression for different baseline severity (figure 15) over up to a two year period of time. In addition, by reconditioning the baseline estimate over a longer period of time, it is possible to estimate a longer time course of an average individual, depending on the starting baseline severity (figure 16).

Model derived estimates obtained from the model for donepezil (where the most complete time course data is available) were compared to those from the Cochrane collaboration review of dementia for the Alzheimer's type (http://www2.cochrane.org/reviews/en/ab001190.html).

In the Cochrane review, 24 trials are included (involving 5796 participants), of which 15 reported results in sufficient detail for the Cochrane meta-analyses. Most trials were of 6 months or less duration. Patients in 20 trials had mild and moderate disease. For cognition there was a statistically significant improvement for both 5 and 10 mg/day of donepezil at 24 weeks compared with placebo on the ADAS-Cog scale (-2.01 points, 95% CI -2.69 to -1.34, P < 0.00001); -2.80 points, 95% CI -3.74 to -2.10, P < 0.00001).

For comparison the model derived estimates over time and the prediction intervals for are shown in figure 17. The model predicted treatment effect is completely in line with that reported for the Cochrane meta-analysis.

Table 9. Model predicted expected mean ADAS-Cog score intercept by baseline MMSE

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 10. Model predicted expected mean change in ADAS-Cog score over one year in the absence of placebo or drug effect, by age

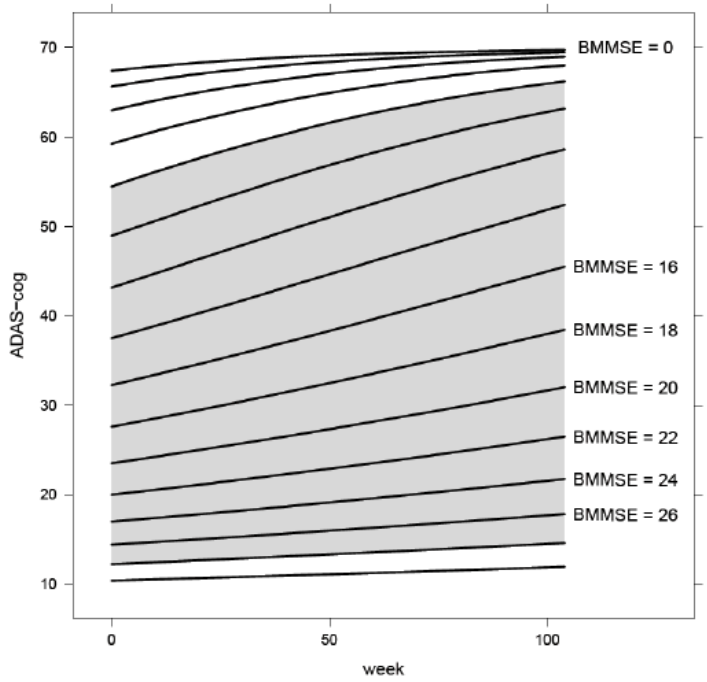
Age	Median	5% LB	95% UB
69	4.92	3.71	6.13
75	4.39	3.51	5.39
80	4.00	2.97	5.17

Table 11. 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

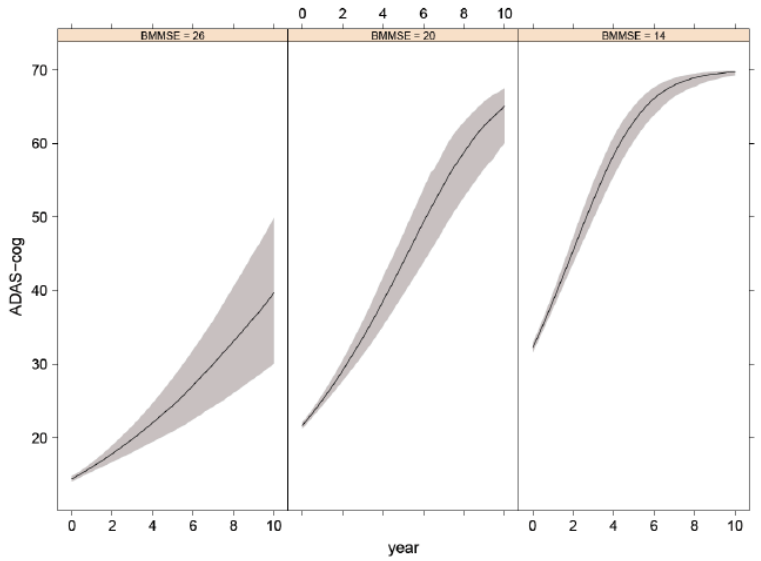
Predicted changes are a function of baseline MMSE, ApoE4 genotype, gender, and age. Age-specific results are not presented because the effects of age and ApoE4 genotype are confounded, preventing independent estimation of their effects (genotype-specific typical age distributions were simulated based on the modeled joint distribution between age and genotype).

Figure 15. Plot of expected 2 year disease progression by baseline MMSE score (average individual)



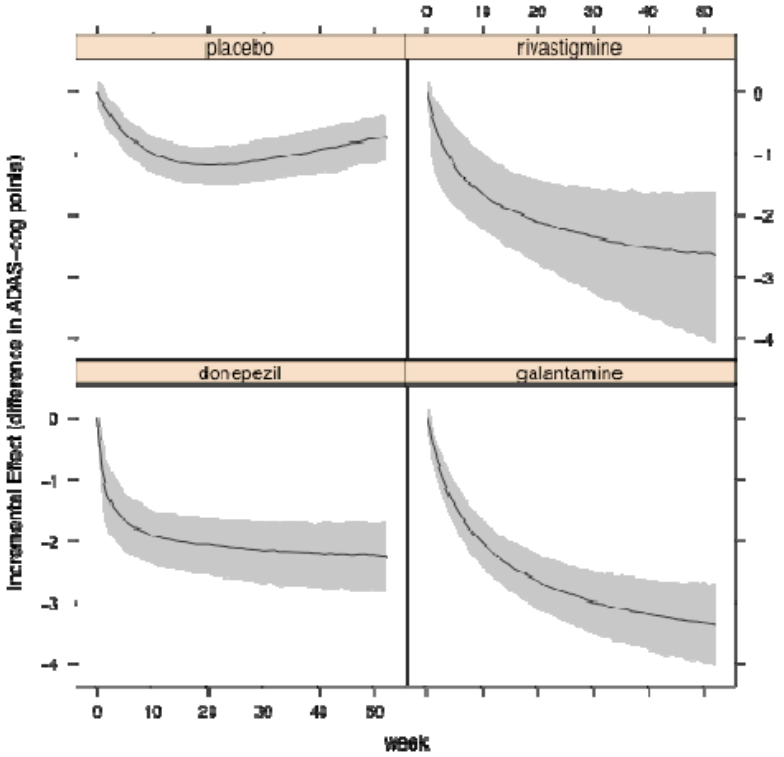
Lines represents posterior median model predictions for a "typical individual" (i.e. with all random effects set to zero) with the given baseline MMSE score. Shaded region the predictions for which there is some support in the data, while predictions outside of the grey region are mathematical extrapolations.

Figure 16. Plot of expected 10 year disease progression by baseline MMSE score (average individual)



Lines represent posterior median predictions for a "typical individual" (i.e. with all random effects set to zero) 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.

Figure 17. Plot of time course for placebo and drug effect model parameters over time



Model posterior median estimates and 90% credible intervals for the incremental effect of placebo (adjusted for natural progression) and for the incremental effects of donepezil, galantamine, and rivastigmine (each adjusted for both natural progression and placebo).

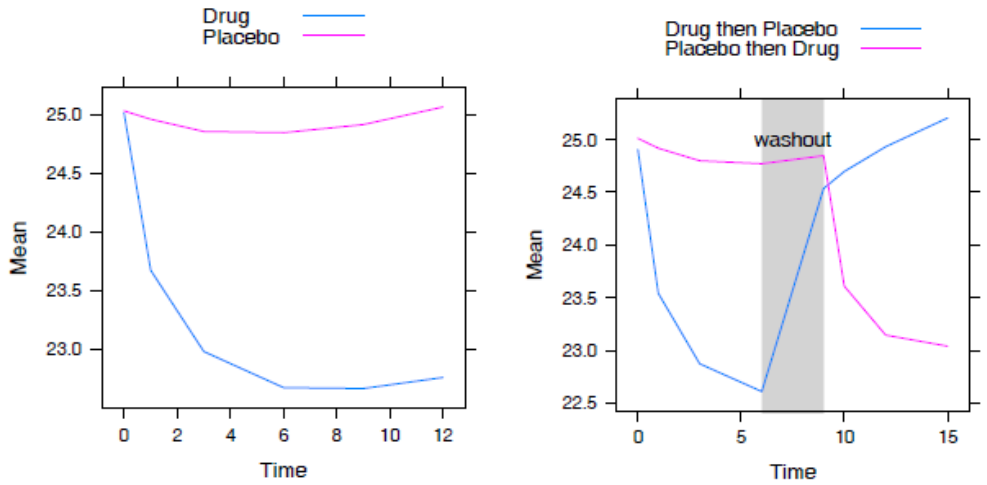
Clinical trial simulations

Symptomatic drug effect scenarios

For compounds where pre-clinical data suggests that only a symptomatic effect is likely to be observed, the key objective in early studies in patients is to determine whether the proposed mechanism translates into meaningful changes in a clinical outcome, as rapidly as possible. Often in early development, duration of toxicology exposure, and concerns for patient safety push teams to explore short, rapid proof of concept (POC) designs. Therefore, exploring the optimal POC studies is of interest.

Figure 18 displays an average simulated trial for 6 week cross-over design (left) and 12 week parallel design (right), respectively under a symptomatic drug scenario. Under these assumptions used for this simulation, treatment effect (difference between placebo and treatment) at the end of each six week period was independent of treatment period in the cross-over design. In this context, a cross-over design has the potential to reduce the sample size while maintaining appropriate power to demonstrate the drug benefit.

Figure 18. Plot of simulated cross-over Trials vs. parallel trials for drugs with only symptomatic effects



In order to have a fixed point of comparison in the evaluation of both designs, the "true effect" of the drug was formulated as the 2.275 point difference at 24 weeks. This corresponds to a common scenario in early phases of drug development in which the estimand of interest is an effect at a time point later than any of the time points studied. From this perspective, some bias is expected with both designs, since the full drug effect is not attained over the duration of the study.

Based on the simulation (table 12), approximately 85% power was achieved with 30 patients per arm (60 patients in total) in a 6-week cross-over study for a symptomatic drug with a drug effect similar to donepezil (2.275 point improvement in ADAS-Cog at 24 weeks, fast drug onset and offset). The power of a 12 week parallel design with 75 patients per arm (150 patients in total) was 77% (table 12). However, as expected, the relative bias of the 6-week treatment in the cross-over study (-11%) was higher than the 12 week parallel study (-4.22%), both of which would underestimate the true steady state treatment effect. A development team may use such results to determine whether the increase in bias is an acceptable price to pay for the gain in power.

Table 12. Comparison of relative bias and power for a 6-week cross-over 12-week parallel study design

Design	Relative Bias (%)	Power
6 week cross-over, n=30/arm	-11.00	0.85
12 week parallel, n=75/arm	-4.22	0.77

Disease modifying drug effect scenarios

For compounds with potential disease modifying effect, 18 month (78 week), randomized, parallel, placebo-controlled trials have most often been selected for use in recent years as summarized by Schneider and Sano. (Schneider and Sano, 2009) As pointed out in the paper, the rationale for their use rests more on historical precedent than objective evidence that this type of trial design would be most sensitive for detecting a disease modifying effect.

An alternative approach, the delayed-start design, has been proposed. This approach can be used to directly support disease modifying claims, based on a series of hierarchical statistical tests of the primary clinical outcomes. Such designs have been implemented in Parkinson's trials. Therefore both designs were simulated for disease modifying drug scenario.

For a 78 week parallel design, the power with 600 patients per group (1200 patients in total) ranged from approximately 17% to 89% with 5% to 50% drug effects on the slope of disease progression (table 13). Anecdotally, these power estimates are somewhat lower than those based on typical design assumptions. This difference is attributable to several factors:

- The model-based estimates for rates of disease progression are generally lower than those used in some power calculations. For example, the model-based estimate for an individual with a baseline MMSE of 21 ranges from approximately 4 to 5.5 points per year, while power calculations have sometimes assumed a rate of progression of 6 points per year.
- The model based estimates of drop-out rates are generally higher than those used in some power calculations. For example, the model estimates approximately 33% drop-out at 78 weeks for a typical mild-to-moderate population, whereas power calculations have sometimes assumed 25% drop-out at 78 weeks.
- The model based estimates of the standard deviation for changes from baseline is higher than that used in some power calculations. For example, the predicted standard deviation for changes from baseline at weeks 26, 52, and 78 are approximately 6, 8, and 10.5 points respectively, whereas power calculations sometimes assume this standard deviation is 8 points for 78 weeks trials.

The power to test the first and second hypothesis in delayed-start design ranged from 8% to 72% when the drug effects on the rate of progression changed from 5% to 50% respectively. As a note, the third hypothesis to test the stability of the treatment difference was not specified and not included in the trial power calculation since no consensus on an appropriate equivalence margin is available for an AD trial. The third hypothesis can be tested later when a clinical meaningful margin is defined. As expected, the power for a 91 week delayed-start design was lower compared to the power of a parallel design for each disease modifying effect assumed. However, the delayed-start design could potentially provide additional inferences for disease modifying effect.

Table 13. Comparison of a 78-week parallel study design and a 91 week delayed-start design by assumption of magnitude of disease modifying effect

Effect	Design	P(reject H_0^1)	P(reject H_0^1 & H_0^2)	H_0^3 5% LB*	H_0^3 95% UB*
5 %	78 week parallel, n=600/arm	0.17			
5 %	91 week delayed start, n=600/arm	0.13	0.079	-0.716	0.685
10 %	78 week parallel, n=600/arm	0.26			
10 %	91 week delayed start, n=600/arm	0.20	0.120	-0.718	0.684
20 %	78 week parallel, n=600/arm	0.53			
20 %	91 week delayed start, n=600/arm	0.44	0.280	-0.742	0.655
30 %	78 week parallel, n=600/arm	0.72			
30 %	91 week delayed start, n=600/arm	0.66	0.460	-0.763	0.637
40 %	78 week parallel, n=600/arm	0.84			
40 %	91 week delayed start, n=600/arm	0.80	0.610	-0.777	0.609
50 %	78 week parallel, n=600/arm	0.89			
50 %	91 week delayed start, n=600/arm	0.88	0.720	-0.794	0.596

* Typical (median) lower and upper bounds for the (treatment-placebo) difference in mean change during the last 6 months of the trial.

Ho 1 No difference in mean ADAS-Cog change from baseline at week 52
Ho 2 No difference in mean ADAS-Cog change from baseline at week 91
Ho 3 Difference in mean ADAS-Cog change from week 65 to week 91 exceeds a given (as yet unspecified) threshold. (Null hypothesis to test non-inferiority, based on treatment-time interaction contrasts.)

Specific questions for EMA review and CAMD positions

DATA

a. Does the Agency agree that the endpoint selected (ADAS-Cog) is suitable for describing cognitive changes in mild and moderate AD?

Applicant's position

ADAS-Cog is a suitable clinical endpoint for describing cognitive changes in patients with mild and moderate AD. Its extensive validation in English and other languages, along with its widespread use over the last two decades, provides the largest and most complete database to describe longitudinal changes in cognition in AD patients. Its value as a measurement tool in clinical trials is further evidenced by the following.

The ADAS-Cog has been the primary cognitive endpoint used for US approvals for all past and currently marketed compounds labeled for the treatment of patients with mild and moderate AD, including tacrine, rivastigmine, and donepezil (note that memantine is indicated for mild and moderate AD, and utilized the severe impairment battery [SIB]).

To CAMD's knowledge, the ADAS-Cog is the agreed primary cognitive endpoint for all recent global phase II and phase III drug development programs in patients with mild and moderate AD. The following late-stage programs all utilized a version of ADAS-Cog (bapineuzumab, ponezumab, solanezumab, Gammagard, Dimebon, SAM-531). As such, ADAS-Cog offers the most value for current and future drug development use. It is acknowledged that the field is moving towards earlier interventions yet the current AD model is established on the wealth of data that exists to date and will serve as a platform for pre-dementia stages as data emerges.

In ongoing natural history studies, such as ADNI and Japanese ADNI (J-ADNI), the ADAS-Cog was the endpoint selected for measuring cognitive change.

All models developed to describe cognition in patients with mild and moderate AD to date have utilized ADAS-Cog, including the work by Faltaos in the AAPS-FDA pharmacometrics fellowship (under the guidance of Dr. Hao Zhu of the Pharmacometrics division of the FDA).

b. Does the Agency agree that the data being used (literature, ADNI, and CAMD database) are sufficient to describe longitudinal changes in ADAS-Cog in patients with mild and moderate AD?

Applicant's position

Both the range and type of data included in the submission is sufficient to describe the longitudinal changes in ADAS-Cog in patients with mild and moderate AD, both for control arms and for treatment arms. CAMD also considers that the data provide sufficient information to inform both drug effect and trial components of the model.

The patient-level control arm dataset which CAMD has been used to support this submission represents data from 3179 patients in 9 interventional trials representing data from all major geographic regions of the world (figure 12). The CAMD dataset which has been used to support this submission has utilized a standard scoring algorithm to ensure cross-study comparability and for potential addition of data in the future.

The range of scores in the dataset includes the entire range of scores from 0 to 70, allowing for validation that simulations at the edges of the distribution of scores are appropriate.

Multiple longitudinal observations in 186 patients with mild AD in the non-interventional ADNI trial represent a reasonable foundation to inform the natural history of AD (table 6).

73 trials in the literature are also included, which provide estimates for the drugs currently marketed for mild and moderate AD. The models by Samtani et al (Samtani et al, 2012) and Faltaos et al (William-Faltaos et al., 2013) do not include data that informs on treatment effects for currently marketed therapies.

The dataset used includes data collected over the last two decades, allowing for temporal comparisons of trends in ADAS-Cog progression over time.

Where available, the dataset includes genotype and biomarker endpoints for testing as covariates. As in the work of Faltaos et al (William-Faltaos et al., 2013) where 581 of the 2479 patients had available ApoE4 status available, not all studies had collected these data.

MODEL

a) Does the Agency agree that that the proposed model provides an adequate quantitative longitudinal description of the progression of cognitive changes in mild and moderate AD for data from various sources? Specifically,

i) That changes in disease progression based on baseline severity have been adequately described?

Applicant's position

Changes in disease progression based on baseline severity are adequately described by the model, as evidenced by the predictive checks from the covariate model, defined as the distribution of the covariate values (the effect of covariates on the response was not defined as part of the covariate model and is described instead as a component of the complete data model). The results are consistent with those reported in previous analyses.

ii) Does the Agency agree that changes in disease progression due to other patient factors (ApoE4 status, gender, age) have been adequately assessed in model development?

Applicant's position

Changes in disease progression due to ApoE4 status, gender and age have been adequately assessed and quantified within the model, as evidenced in table 10, table 11, figure 15, and figure 16.

iii) Does the Agency agree that the internal validation process adequately describes the studies used for model development?

Applicant's position

The progression of ADAS-Cog has been adequately assessed and quantified within the model. Figure 17 illustrates observed versus predicted study-specific standard deviation as a function of time. The fit suggests that the model adequately describes the increases in endpoint variance as a function of time.

v) Does the Agency agree that symptomatic agent effects described by the model for acetylcholinesterase inhibitors are consistent with current clinical opinions?

Applicant's position

Symptomatic agent effects described by the model for AChE inhibitors are consistent with current clinical opinions.

Model derived estimates obtained from the model for donepezil (where the most complete time course data is available) were compared to those from the Cochrane collaboration review of dementia for the Alzheimer's type (http://www2.cochrane.org/reviews/en/ab001190.html).

In the Cochrane review, 24 trials are included (involving 5796 participants), of which 15 reported results in sufficient detail for the Cochrane meta-analyses. Most trials were of 6 months or less duration. Patients in 20 trials had mild and moderate disease. For cognition there was a statistically significant improvement for both 5 and 10 mg/day of donepezil at 24 weeks compared with placebo on the ADAS-Cog scale (-2.01 points, 95% CI -2.69 to -1.34, P < 0.00001); -2.80 points, 95% CI -3.74 to -2.10, P < 0.00001).

For comparison, the model derived estimates and the prediction intervals for the time course for placebo and drug effect model parameters over time are shown in figure 17. The model predicted treatment effect is completely in line with that reported for the Cochrane meta-analysis.

vi) Does the Agency agree that the external validation are sufficient for use for trial simulation purposes?

Applicant's position

The effects described by the model are consistent with current clinical observations. The visual predictive checks provide direct evidence for the goodness of fit (figure 18).

The external validation was conducted for the response to FDA request during the qualification review team meeting on April 28th, 2010. The response data from a randomly selected CAMD protocol (the *test set*) was withheld and blinded from model developers during the model development phase. The fitted model from the *model-building set* 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 (figure 18) to determine if all values fell within the 90% prediction interval.

Based on the criteria, the model met the external validation criteria that had been established. In this case, the baseline ADAS-Cog was estimated to be higher than that observed, placing the observed data in the lower range of what was predicted from the model. In addition the observed change from baseline in this population over 78 weeks was approximately 3 points, less than what would normally be expected, and that which has been observed in other contemporary trials. As such, the model correctly identified the results of this trial as being within the low range of possible outcomes that can be observed in this population.

SIMULATION

a) Does the Agency agree that a simulation approach based on a quantitative model is an adequate strategy for the purpose of comparing clinical trial designs with cognition as a primary endpoint in mild and moderate AD:

Applicant's position

A simulation approach based on a quantitative model is the most suitable strategy for comparing trial designs in mild and moderate AD, where cognition is the primary endpoint.
Simulation techniques can be employed to evaluate the performance of potential designs so that we fully understand the operating characteristics (e.g., probability of false-positive, false–negative, and of making the right decision) of each design based on the current available information about the drug before a specific design is selected. The comparison of the delayed to start to a traditional parallel design provides a direct example of how the model can be utilized in this capacity.

Routine development of trial execution models is recommended so that more quantitative assessment of the impact of protocol deviations can be made. With traditional clinical trial planning, the sample size is set to achieve the desired power, at a selected significance level, assuming a specific fixed treatment effect and variance (often without taking into account changes over time), and perhaps inflated to account for anticipated dropouts. In more informed quantitative drug development, the drug, disease, and trial execution models are used together to predict the treatment effect as a function of dose, regimen, and time.

Moreover, uncertainty in the prediction of the treatment effect can be taken into account from the uncertainty in the parameter estimates of these models. For example, trial-to-trial variation reflecting the uncertainty in the parameters (and indirectly in treatment effect) can be accounted for by parametric or non-parametric bootstrapping techniques. Simulations are then performed using the models together with the bootstrap vectors of parameter values for each simulated trial reflecting the uncertainty, to simulate hypothetical data for each of many simulated clinical trials for each potential design under consideration. Essentially, this approach facilitates the calculation of ''marginal'' power averaged over the uncertainty in the prediction of the treatment effect. This ''marginal'' power calculation leads to a larger sample size relative to assuming the mean treatment effect is known (without uncertainty), but a smaller sample size relative to the worst case one might assume over the range of plausible values given this uncertainty.

If the primary end point involves an imputation method to account for dropout, this is accommodated by simulating time of dropout for each hypothetical subject and applying the imputation method (e.g., last-observation carry forward) to the simulated data. In this case, the drop-out model also contains factors known to influence the dropout, namely baseline severity (figure 15) and age (figure 16). The data analysis is then performed for each simulated trial for each design, and the decision criteria are applied to make a decision (e.g., go or no go). This decision can be compared against the correct decision under the models and true values of the parameter used to simulate each trial. The probability of a correct decision can then be computed as one of the measures of trial performance.

b) Does the Agency agree that the example simulations provided in the submission are sufficient to demonstrate the utility and use of this model as a DDT?

Applicant's position

The use of simulations based on the present model provides an informative strategy for the purpose of comparing the operating characteristics of a wide range of clinical trial design options, using cognition as a primary endpoint in mild-to-moderate AD patients.

The examples provided were selected to illustrate the applicability and usefulness of the tool to help clinical trial design teams compare key operating characteristics of optional designs, including the effect of particular assumptions about the magnitude, onset, and offset of drug effects, varying entry criteria, sample size and design features (i.e. parallel versus crossover) on power, bias and probability of rejecting null hypotheses.

c) Does the Agency agree that this DDT, as defined in this document, is suitable for qualification for use in Drug development as a longitudinal model for describing changes in cognition in patients with mild and moderate AD, and for use in assisting in trial design in mild and moderate AD, as defined by the context of use?

Applicant's position

As defined by the context of use, this DDT ***is suitable to be qualified*** for use in Drug development as a longitudinal model for describing changes in cognition in patients with mild and moderate AD, and for use in assisting in trial design in mild and moderate AD.

The endpoint that is selected the primary endpoint used for cognition in all previous and ongoing studies in mild and moderate AD. CAMD has utilized data from a wide variety of sources including non-interventional natural history, and randomized control interventional studies, spanning the entire range of the ADAS-Cog, and from a broad range of geographical locations.

The model described in this submission represents the current state of the art with respect to a longitudinal disease-drug-trial model to describe changes in ADAS-Cog. The model developed by CAMD built on, and integrated strengths and findings of previously reported models. The present effort focused on issues of estimation, demonstration of model validity, and examples of application. 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 "adopted" features included:

1. The use of a generalized logistic function to describe the natural progression of the disease on a constrained scale (Samtani et al., 2012).
2. The use of a Bateman-type function to describe the incremental placebo effect (Holford and Peace 1992, Ito et al., 2010).
3. The use of Emax functions to describe the incremental effects of approved AChE inhibitors as a function of dose (Ito et al., 2010, Gillespie, 2009).
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 (Ito et al., 2010, Samtani et al., 2012).

5. The use of baseline age and baseline severity as covariates on the hazard of drop-out (William-Faltaos et al., 2013).

In addition, CAMD has incorporated a number of important innovations:

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. The use of the generalized logistic function is itself only sufficient to ensure that conditional expectations respect the 0-70 constraints. However, when the generalized logistic function is used in conjunction with normally distributed residuals, there is a positive probability of simulating results outside of the allowable range. The Beta-distribution eliminates this.

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).

Based on the coordinators’ report, the Scientific Advice Working Party held that before opinion can be provided the applicant should discuss the following points:

Summary

The tool is a clinical trial simulation tool to help optimize clinical trial design for mild and moderate AD, using ADAS-cog as the primary cognitive endpoint. It is based on a drug-disease-trial model that describes disease progression, drug effects, dropout rates, placebo effect, and relevant sources of variability. It is not intended for the approval of medical products without the actual execution of well conducted trials in real patients.

Scientific discussion

This seems to be a useful approach to enable better and more informed decisions to be made during the process of designing trials in Alzheimer’s disease. From a CHMP perspective, the simulation tool is not intended to replace clinical trial data so in the end there will always be a phase III trial on which to base the assessment. In this context it is easy to welcome the availability of such a tool, though it is still important to ensure the simulations lead to good design solutions.

The model was fitted based upon the CAMD data-base consisting of 9 trials with 3223 patients. These data were used to inform about inter-patient variability, patient specific factors and placebo effect. The group also looked at 73 trials from the literature to inform about inter-study variability and the effects of marketed drugs. After fitting the model was validated using placebo data from study 1014, which wasn’t included in the model fitting. This study included 639 subjects and the fit appeared fairly good, though it would be interesting to see more details.

It should be clarified if this was the full extent of the independent validation, or if further work was done. If not it would still be interesting to see how good the fit is to the trials that were included in the model selection. There would be interest in how sensitive the fitted model is to the choice of data-set. For example how much would the fitted model alter for different choices of fitting/validation set. Are there any plans to continually validate/reassess the model as further trials become available?

The Applicant presented the fitted rates of progression as a function of baseline factors. The baseline factors included in the ADAS-cog model were mini mental state examination score (MMSE), APOE4 status, age and gender. It seems that baseline ADAS-Cog was not included as a covariate. This would be expected to be highly predictive, though is maybe correlated with MMSE. The results showed that estimated progression was faster for males than females, faster for those with lower baseline MMSE score (low scores indicate worse cognition). The pattern wasn't as clear for baseline APOE4 status. While those with 2 alleles had faster progression, there was little difference between those with 0 or 1. Baseline age was also used as a covariate, but progression rates weren't presented as a function of age.

It is acknowledged that the model is for mild-moderate Alzheimer's disease, but it would be interesting to know if the model has any validity if extrapolated outside this range i.e. to prodromal and severe disease.

Overall this approach seems to have the potential to be a valuable tool to improve the design of clinical trials. As an illustration of the benefit it would be interesting to see a hypothetical parallel group trial for a symptomatic treatment powered conventionally and using the simulation tool, and see what difference it could potentially make to the patient number/design decisions.

List of issues addressed during the discussion meeting

SAWP/CHMP question
Can you describe the process used for selecting the covariates for the model and what other factors were considered aside from those included? In particular was baseline ADAS-Cog a strong contender for inclusion?

Applicant's position

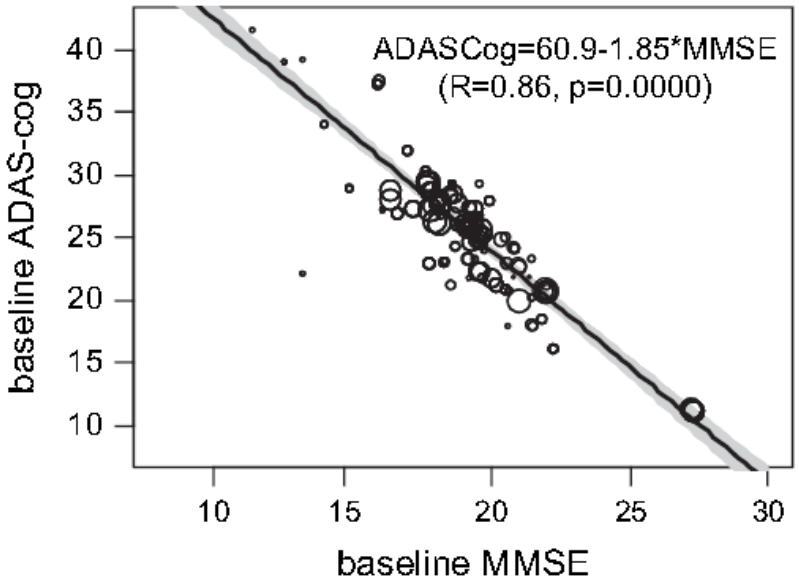
Baseline severity is a strong predictor for disease progression, which is consistent across previously published results [Ito et al[1,2], Samtani et al[3], William-Faltos[4]] and the CAMD analysis. Previous results reported ADAS-Cog as an indicator as baseline severity, and was included in their models. For CAMD model, however, MMSE was used instead of ADAS-Cog as an indicator for baseline severity, given that 1) high correlation with ADAS-Cog (r=0.86), 2) MMSE is the most widely used assessment for screening purpose, which also serve as a useful measurement for clinical study simulations.
As described in section 2.3.3.3 of the submission document, the covariate selection process was based on previous work by Ito et al[1], Samtani et al[2], evaluated baseline age, ApoOE4 genotype, family history of AD, gender, years of education, and baseline MMSE. In Ito's previous work, continuous variables (age, education, baseline ADAS-Cog) were normalized to a value representative of the population for that variable, that is, the approximate mean value of the dataset. For baseline severity, Ito et al[2]., tested an inverse-U type function (modified inverse-U function) in addition to the power function, to describe the nonlinear relationship between the rate of change (slope) and severity (baseline ADAS-Cog score). In turn, Samtani et al[3], tested an initial list of 34 covariates. These covariates fell into the following four categories: 1) MRI volumetric measures 2) serum biomarkers 3) demographic and genetic factors and 4) cognitive tests at baseline/screening. After excluding correlated covariates, or creating single summary variable was created to represent correlated

predictors previously identified as important (using an absolute correlation coefficient value r > 0.3 as cutoff), the following relevant covariates (on disease progression) were selected: APOε4, total serum cholesterol and age. With this in mind, the CAMD team took advantage of this previously relevant work for covariate selection purposes, with the following modifications:

i. Education (as tested by Ito et al[1], and Samtani et al[3]) was not included given high variability in the way these data were captured in the CAMD database, which escaped even the CDISC standardization process.

ii. MRI volumetric measures, CSF biomarkers and total cholesterol were not included since they were not consistently represented across the trials available in the CAMD database.

iii. Previous work from Ito et al, found baseline ADAS-Cog and baseline MMSE to be correlated (see figure below from Ito et al[1]). As discussed in section 2.3.3.3, from a trial simulation perspective, it is preferable to develop a model in which all covariates represent potential trial entry criteria, since altering these variables allows the clinical trial design team to directly observe the impact they have on trial design. Whereas the MMSE is designed as a screening tool and is almost universally incorporated in inclusion/exclusion criteria, the duration of time needed to administer the ADAS-Cog renders this instrument far less practical for screening purposes and hence less useful as a model covariate.



Correlation between baseline ADAS-Cog and baseline MMSE

SAWP answer

SAWP asked if baseline ADAS-Cog had actually been tested as a covariate and compared against the selected baseline MMSE covariate. The Applicant responded that baseline ADAS-Cog was not actually tested as a covariate for the reasons described above, which is considered acceptable. The Applicant further explained that if one were to perform analyses using baseline ADAS-Cog as a covariate when the outcome measure is ADAS-Cog it would not represent an independent measure. The CHMP guideline on adjustment for baseline covariates, states verbatim, "If a baseline value of a continuous outcome measure is available, then this should usually be included as a covariate. This applies whether the primary outcome variable is defined as the 'raw outcome' or 'change from baseline.'"[5] SAWP acknowledged that although it would have been interesting to include baseline ADAS-Cog as a

covariate, the presented correlation between baseline MMSE and baseline ADAS-Cog, would likely yield similar results.

SAWP/CHMP question

Can you clarify the scale used when describing progression rates from the fitted model? Is it points per year on the ADAS-Cog?

Applicant's position

Applicant said yes, the progression rate represents points per year on the ADAS-Cog. Applicant pointed out that in developing the CAMD database, there was a requirement to remap ADAS-Cog to a common standard despite the fact that most experts assumed it was uniform across clinical trials. Even in the presence of various versions, (11, 13, 14 etc.), the uniform ADAS-Cog 11 could be extracted from all the trials available for analysis.

SAWP answer

It was explained that ADAS-Cog11 was used for the model.

SAWP/CHMP question

The model suggests there is little difference in progression rates between those with APOe4 status 0 and 1. Is this a plausible finding?

Applicant's position

As described in Table 11 of the submission document, homozygous APOe4 carriers have a clearly higher progression rate against comparable individuals from the same gender and with equivalent baseline severity. Due to sample size limitations, it was not feasible to more thoroughly evaluate the effects of all possible allele combinations, especially when such combinations included potential protective effects from the other allele variants such as APOe2. Applicant expressed that this was done based on feedback received from FDA, based on the rationale of trying to better characterize potential risk differences between the heterozygous and homozygous carriers.

SAWP answer

SAWP asked the reason why three categories of APOe4 status were defined, as opposed to the carrier/non-carrier binary conversion more frequently used. SAWP suggested the possibility of compressing APOe4 status into two groups, given the little added information differentiating status 0 and 1.

Applicant's question

Please describe how the estimated progression rate varies with baseline age.

Applicant's position

Tool Estimates Rates of Progression as a Function of Clinically Important Baseline Factors

Yearly Rate of Change in ADAS-Cog

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

CAMD CONFIDENTIAL 22

As described on Table, younger individuals at baseline showed a higher rate of progression (almost a full point per year when comparing 69 year-old with 80 year-old individuals).

SAWP answer

The SAWP agrees that the estimate rates of progression can help in the design of the studies.

SAWP/CHMP question

Please clarify whether the applicant in their modelling exercise has considered any functional outcomes and whether they were correlated with changes in ADAS-Cog?

Applicant's Position

Functional endpoints were initially considered for inclusion in the analyses. However, as opposed to the ADAS-Cog 11, which was consistently collected in all the trials in the CAMD database, ADNI, and in the summary literature reports included in the metadata, the functional endpoints included varied greatly. From a practical perspective, and based on feedback received from FDA and EMA, the CAMD team decided to focus on the ADAS-Cog11 as the modeling endpoint.

Previous work with AChE inhibitors like that of Rogers et al[5], has shown that ADAS-Cog and functional endpoints such as the Clinician's Interview−Based Impression of Change including caregiver information (CIBIC plus) are indeed correlated.

SAWP answer

The applicant showed results illustrating correlation between ADAS-Cog and specific functional measures in ADNI and other data sources in mild/moderate AD cases. The applicant claims that the ADAS-Cog was chosen and the endpoint for the model given it is the most consistently used whereas different studies tend to use different functional scales/measures. Some work by team members has analyzed model predictions in subsets of studies using specific functional endpoints (FAQ) and the results are similar to that observed with ADAS-Cog. The current CHMP guidelines for Dementia include ADAS-Cog as a primary end point for efficacy. Thus, SAWP is satisfied that is a clinical meaningful measure.

SAWP/CHMP question

A question was raised if there should be a model for both cognition and function. The answer is yes, in the future. Consensus on functional endpoints to be implemented in trials would accelerate progress in achieving this goal in the future. Please provide further details of the results of the independent validation using study 1014.

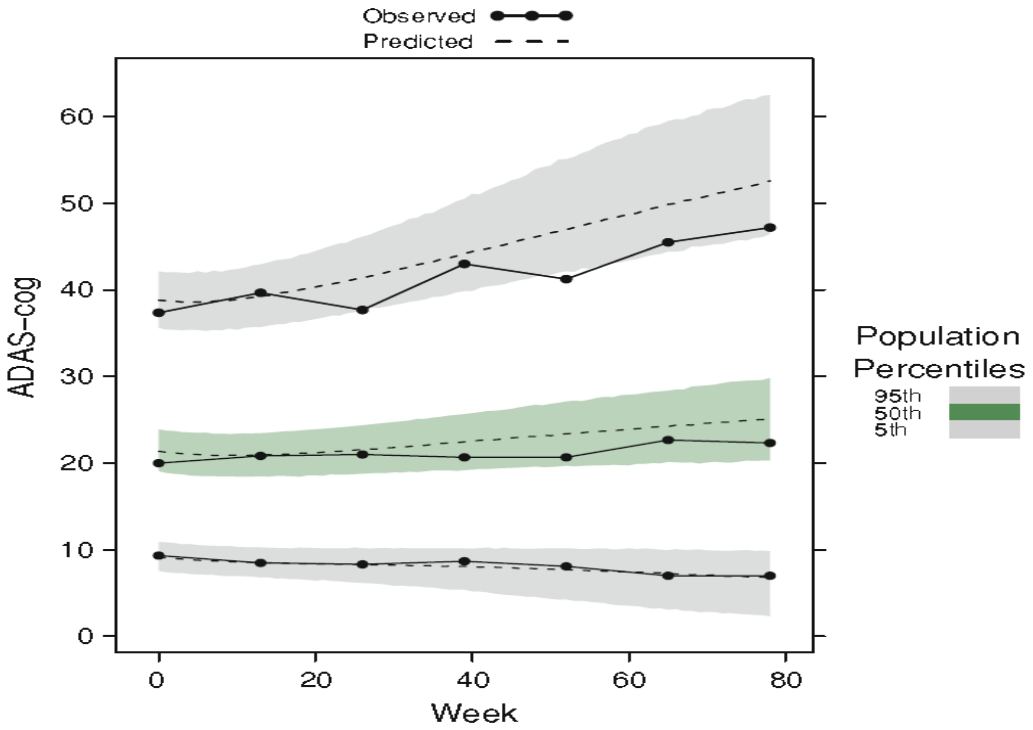
Applicant's position

The response data from study 1014 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 (not conditioned on study-level random effects). The predictive validity of the model was then assessed by graphically comparing the observed data to the model predictions (see figure below) to determine if all values fell within the 90% prediction interval. A discussion regarding external validation included description of study 1014 external validation, consistency with published literature and findings from others (Holford and Peace[6], Schneider and Sano[7], Samtani[3]).

Applicant pointed out that 4 of the seven studies in the CAMD database included treatment with stable background therapy. The other 3 studies were placebo only. No differences in the rate of disease progression were observed based upon background therapies. Interest in combination therapies in this patient population was highlighted and a question for the future may relate to understanding how much background treatment is considered relevant.

SAWP answer

The SAWP is of the view that the value of the model is not only based on the studies used to develop the model. The model was derived not only from the CAMD placebo database but also ADNI and 83 studies from the published literature. It is of relevance that placebo response was derived from the CAMD database, symptomatic response from literature and disease progression from ADNI. Thus, the model novelty is in being derived from a comprehensive diverse integration of data.



Unconditional predictive check for external validation

SAWP/CHMP question

Was there any additional independent validation aside from study 1014?

Applicant's position

Additional independent validation steps in *stricto sensu* were not carried out.

SAWP answer

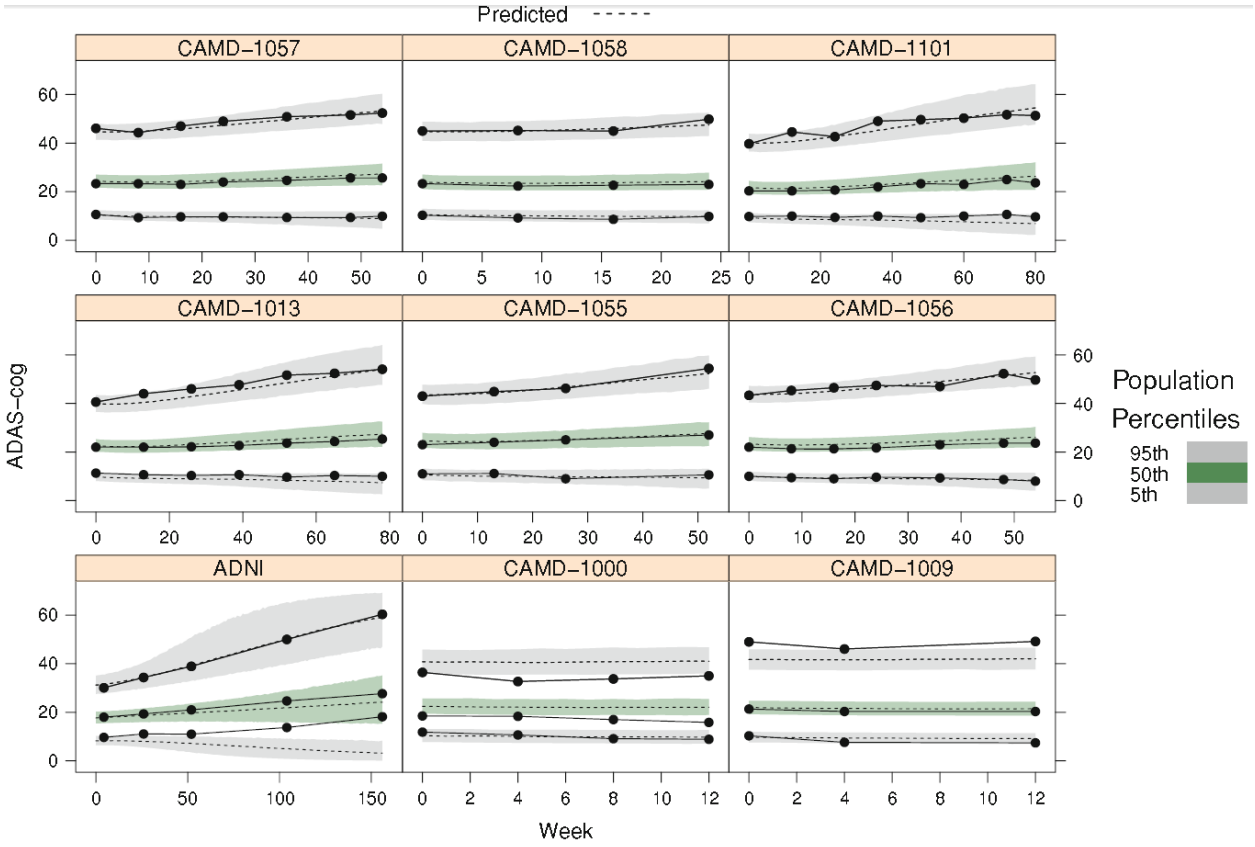
The SAWP has recommended further working with industry partners to run the model against datasets not previously used. The applicant stated that several similar predictive distributions for such data are being performed in the industry with similar results.

SAWP/CHMP question

Can you show how well the model fits each of the studies that were included in the fitting?

Applicant's position

The figures below illustrate the unconditional predictive checks for sample population percentiles of ADNI and CAMD studies.



Unconditional predictive checks for internal validation

SAWP answer

The example of the unconditional predictive checks for internal validation was relevant for sample population.

SAWP/CHMP question

How robust is the fitted model to the choice of fitting/validation sets? For example if each study was removed in turn to be the validation set, with the other n-1 being used for fitting, how much would the model alter and would the validation still look good?

Applicant's position

The validation dataset (study 1014) was randomly selected from the available dataset in CAMD database (which met the selection criteria: ≥1 year & ≥100 patients) before starting the model building process. Also, the applicant compared the final parameter estimates with/without 1014 after completion of the external validation, and we didn't see any outstanding difference. Therefore, as long as the dataset meets the selection criteria, we believe the validation using other dataset would be similar with what we demonstrated in the submission document.

Additionally, since the modeling strategy did not involve any substantial variable selection (only relatively few covariates were available, and their role in the model was largely pre-specified), it may be reasonably expected that leave-one-out cross-validation (as referred to in the question) would produce results extremely similar to those seen with the posterior predictive checks. One generally only finds disagreement between cross-validation and posterior predictive checks when the modeling strategy involves substantial variable selection. CAMD explained how robust is the fitted model with

description of jackknife approach, comparison of final parameter estimates and model prediction yielding similar results with or without study 1014.

SAWP answer

The model is robust to fulfill the choice of validation sets.

SAWP/CHMP question
Are there plans to continually update/validate the model as new data becomes available?

Applicant's position

The CAMD team envisions modeling and simulation tools as continuously evolving entities that should be in a constant process of enrichment, refinement and expansion. Examples such as integration of biomarkers into the model were highlighted. However, it is important to note the essential role of precompetitive data sharing and magnitude of effort and resources required to remap additional datasets and perform the data QC process in order to expand the CAMD database.

SAWP answer

The SAWP recommends that the modeling and simulation tools will be continuously evolving entities that should be in a constant process of enrichment, refinement and expansion.

SAWP/CHMP question
Does the model have any validity if extrapolated outside the mild/moderate range, i.e. to prodromal or severe disease?

Applicant's position

A preliminary extrapolation into the more severe states is shown on figure 25 of the submission document, in which the predictive progression curves for 65 year-old ApoE4 non-carrier males is shown over a ten-year period. One caveat, though, is the potential limitations of the ADAS-Cog as an outcome measure in a more severe population, where scores would tend to compress against the maximum 70 points of this scale.

Conversely, extrapolations into prodromal or pre-demented stages have not been attempted, mainly due to the limited amounts of clinical trial data in these populations, and the potential limitations of the ADAS-Cog as an outcome measure in such stages of disease.

Finally, as stated in the proposed context of use statement, the model is intended for application in the mild and moderate AD stages, not in pre-dementia or severe dementia stages.

SAWP answer

The SAWP agrees that the model can be used for design of trials in mild and moderate AD, not in pre-dementia or severe dementia stages.

SAWP/CHMP question
Can you provide a hypothetical example showing how a basic trial might be powered both with and without the simulation tool?

Applicant's position

As illustrated in the example on section 2.4.6.2 of the submission document, a development team might find themselves confronted with designing a trial for a drug to evaluate a drug with a potential

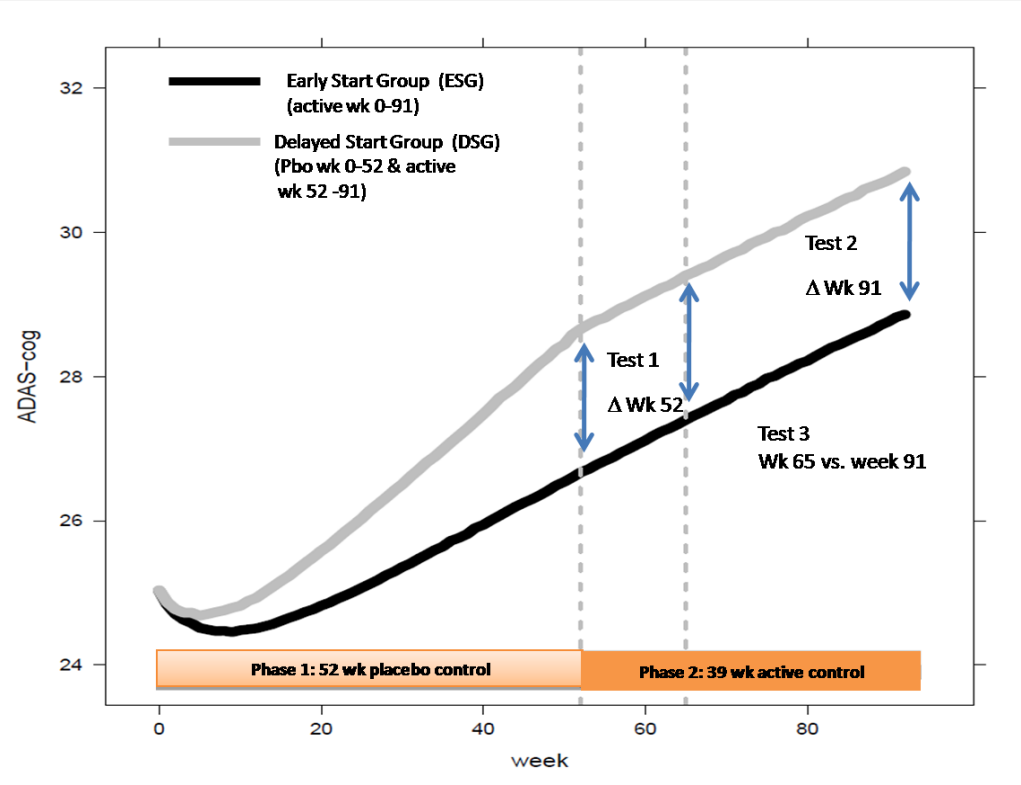
disease modifying effect. For compounds with potential disease modifying effect, 18 month (78 week), randomized, parallel, placebo-controlled trials have most often been selected for use in recent years as summarized by Schneider and Sano[6]. As these authors point out, the rationale for their use rests more on historical precedent than objective evidence that this type of trial design would be most sensitive for detecting a disease modifying effect. An alternative approach, the delayed-start design, has been proposed. This approach can be used to directly support disease modifying claims, based on a series of hierarchical statistical tests of the primary clinical outcomes. Such designs have been implemented in Parkinson's trials. Therefore both designs were simulated for disease modifying drug scenario described in section 2.3.3.9 of the submission document. A team confronted with these two design options could base the clinical trial design process on selecting one of the two designs without much quantitative background, other than the historical frequency of use that would support selecting a parallel design, versus the extrapolation from Parkinson's disease trials that have used the delayed start framework. In either case, the team would need to then have the option of basing the expected progression rates on the opinion of clinical experts, or develop their own model-based understanding regarding expected progression rates based on in-house data available to them.

The former option would likely provide a one-size-fits-all estimate of progression rates, without much consideration for varying progression rates in subpopulations defined by relevant covariates, with model-based estimates for rates of disease progression generally being lower than those used in some power calculations. (For example, the model-based estimate for an individual with a baseline MMSE of 21 ranges from approximately 4 to 5.5 points per year, while power calculations have sometimes assumed a rate of progression of 6 points per year.) The latter option could be an interim solution, but would potentially lack the level of underlying data, while the case of CAMD is based on a large scale patient-level and summary-level integration of data likely without precedent in the field of Alzheimer's disease. A similar scenario would present itself in the case for expected sample size attrition rates. The model based estimates of drop-out rates are generally higher than those used in some power calculations. For example, the model estimates approximately 33% drop-out at 78 weeks for a typical mild-to-moderate population, whereas power calculations have sometimes assumed 25% drop-out at 78 weeks. Additionally, the model based estimates of the standard deviation for changes from baseline is higher than that used in some power calculations. For example, the predicted standard deviation for changes from baseline at weeks 26, 52, and 78 are approximately 6, 8, and 10.5 points respectively, whereas power calculations sometimes assume this standard deviation is 8 points for 78 weeks trials.

If, on the other hand, the team decided to make use of the proposed clinical trial simulation tool, varying disease progression rates could be generated based on a range of entry criteria variations (as opposed to a one-size-fits-all approach). As illustrated on sections 2.4.6.2 of the submission document, disease modifying drug effects were expressed as proportional reductions in expected progression rates. Based on the feasibility to detect a potential effect, the proportional reductions considered were 20%, 30%, 40%, and 50%. The sample size for simulated trials included 100, 250, 400 and 600 per group. The candidate designs considered in these scenarios were: A two-arm parallel design with 78 week treatment duration and assessments at weeks 0, 26, 52, and 78. The assumed primary analysis used a Multivariate Model for Repeated Measures (MMRM) approach with unstructured covariance matrix and fixed effects for baseline ADAS-Cog, treatment, visit (nominal), and treatment by visit interaction. Drug effect was formulated as the expected difference at week 78.

A two-group delayed start design. This design employs a placebo-control stage (stage 1), and an active control phase (stage 2). The patients who receive placebo in the placebo control phase and study drug in the active control phase are referred to as the delayed-start group. The patients who receive study drug in both phases are referred to as the early-start group. 52 week and 39 week duration was selected for stage 1 and stage 2, respectively, with the final 26 weeks being used to

assess stability of effect. Assessments were assumed at weeks 0, 26, 52, 65, 78, and 91. A schematic for this design is provided in the figure below.



Schematic of the delayed-start design

The envisioned primary analysis would test the three research hypotheses associated with delayed start designs:

i. Test for the difference in ADAS-Cog change from baseline between the placebo and study drug group at the end of phase 1 (52 week).

ii. Test for the difference in ADAS-Cog change from baseline between early and delayed start groups at the end of phase 2 (91 week).

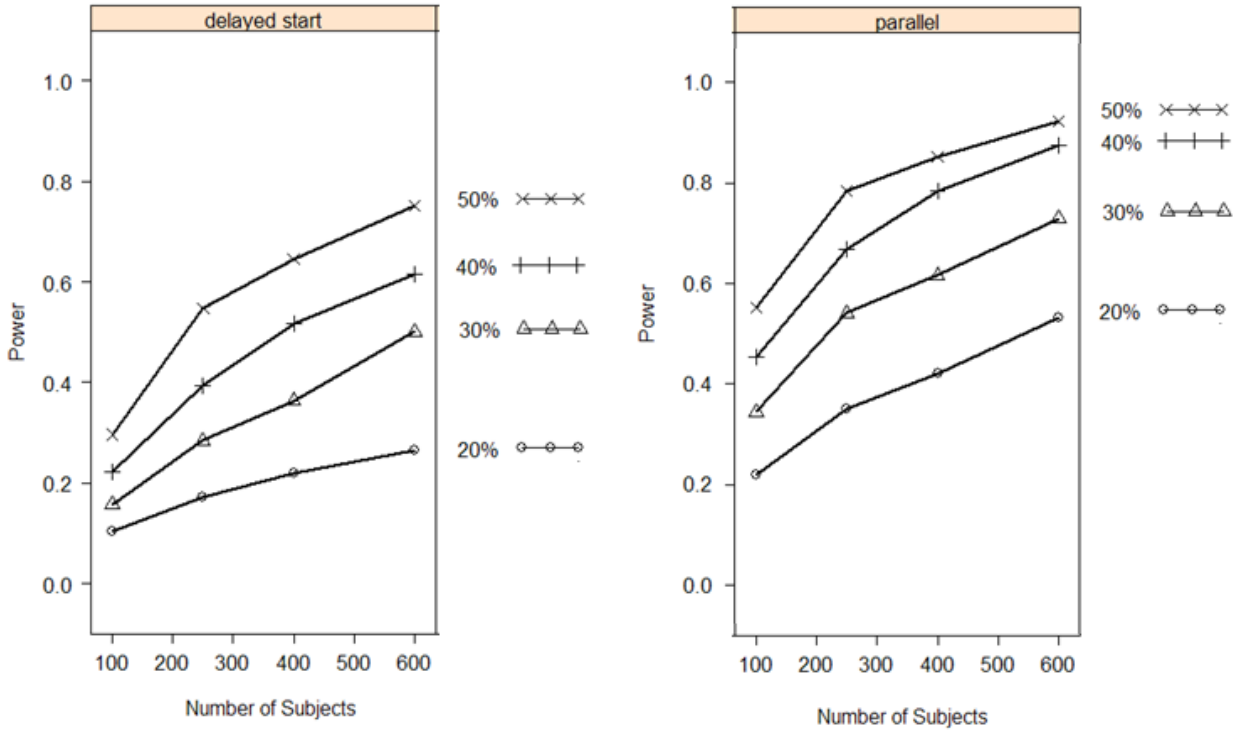
iii. Test for evidence of the stability of the treatment difference, which may be assessed by comparing the change from week 65 to week 91 for early versus delayed start groups.

A formulation of any of these three hypotheses in terms of slopes is possible in general, but would be conceptually inconsistent with our present model, which implies non-linearity of the time courses. Moreover, in the ADAGIO study (the delayed start trial in Parkinson's disease for rasagiline), the slope analysis was pre-specified and used for hypothesis 1 and 3 testing but the data failed the non-linearity tests and as a result, the slope tests were considered inconclusive (http://www.fda.gov/downloads/AdvisoryCommittees/CommitteesMeetingMaterials/Drugs/PeripheralandCentralNervousSystemDrugsAdvisoryCommittee/UCM277005.pdf). Consequently, all three hypotheses were tested using interaction contrasts rather than slopes, using the same MMRM model as described for the 78 week parallel design described before. Also since there is no consensus regarding an appropriate equivalence margin for testing the stability of effect (whether formulated as a slope or an interaction contrast), the third hypothesis was not included in the trial power calculation.

For the 18 month parallel design, approximately 85% power was achieved with 600 patients and 400 patients per group for decreases of 40 and 50% on the rate of disease progression, respectively (see

figure below). The power to reject both the first and second hypothesis in the delayed-start design was much lower compared to the parallel design (see figure below). For a moderate disease modifying effect of 50% decrease on the rate of disease progression, approximately 75% power was achieved with 600 patients, although the delayed-start design could potentially provide additional inference for a disease modifying effect.

As can be seen, the power to test the first and second hypothesis in delayed-start design ranged from 8% to 72% when the drug effects on the rate of progression changed from 5% to 50% respectively. As indicated above, the third hypothesis to test the stability of the treatment difference was not specified and not included in the trial power calculation since no consensus on an appropriate equivalence margin is available for an AD trial. However, this third hypothesis can be tested later, once a clinical meaningful margin is defined (either through consensus in the literature or through feedback from the regulatory agencies during the interactions between the sponsor and regulators). As expected, the power for a 91 week delayed-start design was lower compared to the power of a parallel design for each disease modifying effect assumed. However, the delayed-start design could potentially provide additional inferences for disease modifying effect.



Power curve of a 78 week parallel study design and a 91 week delayed-start design by assumption of different magnitude of disease modifying effect

SAWP answer

The applicant showed several examples such as:

a) The model could be used to power a clinical trial being designed prospectively.

b) Highlighted ability to accurately predict drug effects, dropout rates, assess disease modifying vs. symptomatic or even combined mechanisms of action based on what is known about the drug in nonclinical studies.

c) The model could be used to predict how fast a drug response is expected to be observed, and to conduct 'what if' scenarios based on defined covariates.

The impact of the model in specific clinical trial designs was discussed including parallel designs, crossover and delayed start designs. One example that generated interest is the use of the model to conduct futility analysis with confidence in deciding if it makes sense to make a go/no go decision on advancing a candidate further in clinical development. Another example was post hoc analyses to look for subsets of patients that respond to treatment to justify further support for new trials in subpopulations. The Applicant highlighted the >90% failure rate in AD trials which served to emphasize the impact of how disease modeling can be implemented in the future to reduce the risk of failure due to poor trial design or other such factors. The SAWP agrees that the model can help to improve efficiencies in relation to the scope of the model.

The SAWP recommends that the model will be made publically available for free, and that the CAMD modeling team along with Metrum research group will assist with training for those who have interest.

Based on the qualification team report the CHMP gave the following answers:

Qualification of a novel data driven model of disease progression and trial evaluation in mild and moderate Alzheimer 's disease.

Context of use

The context of use: "The proposed disease progression and trial evaluation model, as defined in this document, *is suitable for qualification* for use in drug development as a longitudinal model for describing changes in cognition in patients with mild and moderate AD, and for use in trial designs in mild and moderate AD."

CHMP Qualification opinion

The proposed disease progression and trial evaluation model, as defined in this document, ***is suitable for qualification*** for use in drug development as a longitudinal model for describing changes in cognition in patients with mild and moderate AD, and for use in assisting in trial designs in mild and moderate AD, as defined by the context of use.

It is important to note that there is no intention to use the model as a replacement for clinical trial data, and such an initiative would not be supported. Appropriate internal control arms, including use of placebo, should continue to be used in prospective randomized controlled trials. The model is also not intended to replace scientific judgment over interpretation of clinical data and/or guidance over clinical drug development. The results of post-hoc analyses would still need to be treated with the usual caution. However, use of the model may help a sponsor to elucidate their level of belief in a hypothesis generated post-hoc to help decide whether to perform a trial to confirm that hypothesis or not pursue it further.

Having such a quantitative framework does not preclude that a given sponsor may use other quantitative tools to support decision-making during the clinical trial design process, but provides valuable information to improve decision making and a unique common backdrop to facilitate quantitative-based discussions between sponsors and regulators.

Also, as acknowledged by the applicant, the model is specifically tailored for mild and moderate AD and has no validity outside this range, e.g. severe or prodromal Alzheimer's disease.

CAMD has provided several clinical trial simulations that were run for illustrative purposes. These simulations were not intended to provide evidence toward any global preference of a particular design, but as examples of how a development team might use the model and associated simulation tools to select designs that are tailored to particular assumptions about the magnitude, onset, and offset of drug effects.

ADAS-Cog is the primary endpoint used for cognition in all previous and ongoing studies in mild and moderate AD. CAMD has selected data from a wide variety of sources including non-interventional natural history, and randomized control interventional studies, spanning the entire range of the ADAS-Cog, and from a broad range of geographical locations.

The model developed by CAMD built on and integrated strengths of previously reported models; the model provides satisfactory information to support its use for simulation. The model can be used to simulate the natural progression of disease (without placebo or drug effect), the progression on a placebo arm, or on a drug arm (either symptomatic or disease modifying). Simulations can be used to inform on the power of competing designs for a clinical trial by simulating data from a placebo arm and from an active arm based on an assumption about the “true” size of benefit.

The choice of covariates for the model was limited by the data available in the studies being used by the modeling, but for those that were included (baseline MMSE, ApoE4 status, gender, age) the fitted relationship to ADAS-Cog is both clinically plausible and a good fit to the data. Functions to model the placebo effect and active arm effects are also included.

In terms of internal validation, the model is a good fit to the data from the large majority of the studies used in developing the model. Using the example of donepezil it has also been demonstrated that the behavior over time of patients on a symptomatic treatment arm can be modeled. The results were consistent with those seen in the Cochrane data-base.

In an external validation exercise the model provided satisfactory predictions of a data-set that had not been included in the modeling exercise, providing reassurance that simulations from the model can be informative regarding the likely changes in cognition of patient groups in clinical studies.

There is some caution to be expressed on the applicability of the model. The model is necessarily built based, in part, on existing clinical trial data which recruited a particular type of patient based on the various inclusion and exclusion criteria and based on the judgment of patients, caregivers and physicians on the perceived suitability of a particular trial for a particular patient. As the patient population changes over time, or as patient management or the natural course of the disease change over time (in ways not necessarily captured by the factors included in the model), the applicability of the model would need to be verified. It is also the case that trials may be conducted using a patient population that is enriched for a particular characteristic. There is also caution that assessment of cognition may be made differently in trials with an active control arm rather than trials with a placebo control arm. These represent reasons to encourage continual development and validation of the model.

The extensive efforts undertaken to build and validate the model are recognized. Further work is encouraged to integrate information on disease progression according to biomarker profiling and to extend the range of the model (or another model) into prodromal AD. Of course, an assessment of function is of clinical and regulatory interest in addition to the assessment of cognition.

The response given by CHMP is based on the questions and supporting documentation submitted by the Applicant, considered in the light of the current state-of-the-art in the relevant scientific fields.

References

Ahn JE, French JL. Longitudinal aggregate data model-based meta-analysis with NONMEM: approaches to handling within treatment arm correlation. J Pharmacokinet Phar 2010;37(2):179-201.

Atchison TB, Massman PJ, Doody RS. Baseline cognitive function predicts rate of decline in basic-care abilities of individuals with dementia of the Alzheimer's type. Arch Clin Neuropsych 2007; 22(1):99-107.

Bonate PL. Clinical trial simulation in drug development. Pharm Res 2000; 17(3):252-6.

CAMD AD Modeling and Simulation Sub Team Meeting Notes (July 7, 2011).

Chan PLS, Holford NHG. Drug treatment effects on disease progression. Annu Rev Pharmacol 2001; 41:625–59.

D'Agostino RB. The delayed-start study design. New Engl J Med 2009;361(13):1304-6.

Gillespie, W. Population dose-response model for ADAS-Cog scores in patients with Alzheimer's disease by meta-analysis of a mixture of summary and individual data. American Conference on Pharmacometrics, Mashantucket, CT, October 4-7, 2009.

Gobburu JVS, Lesko LJ. Quantitative disease, drug, and trial models. Annu Rev Pharmacol 2009; 49:291-301.

Hennig S, Nyberg J, Hooker AC, et al. Trial treatment length optimization with an emphasis on disease progression studies. J Clin Pharmacol 2009; 49: 323-35.

Holford NHG, Kimko HC, Monteleone JPR, et al. Simulation of clinical trials. Annu Rev Pharmacol 2000; 40:209-34.

Holford NHG, Peace KE. Methodologic aspect of a population pharmacodynamic model for cognitive effects in Alzheimer's patients treated with tacrine. P Natl Acad Sci USA 1992;89: 11466-70.

Ito K, Ahadieh S, Corrigan B, et al. Disease progression meta-analysis model in Alzheimer's disease. Alzheimers Dement 2010;6(1):39-53.

Ito K, Corrigan B, Zhao Q, et al. Disease progression model for cognitive deterioration from Alzheimer's Disease Neuroimaging Initiative database. Alzheimers Dement. 2011;7(2):151-60.

Ito K, Romero K, Neville J, Rogers JA, Corrigan B. Understanding Placebo Responses in Alzheimer's Disease Clinical trials from the literature meta-data and CAMD database. ASCPT, March 2012.

Lambert PC, Sutton AJ, Burton PR, et al. How vague is vague? A simulation study of the impact of the use of vague prior distributions in MCMC using WinBUGS. Stat Med 2005;24(15):2401-28.

L'Ecuyer P, Simard R, Chen EJ, et al. An object-oriented random-number package with many long streams and substreams. Oper Res 2002;50(6):1073-5.

Little RJA, Rubin DB. Statistical analysis with missing data. 2nd ed: Wiley; September 2002.

Lockwood P, Ewy W, Hermann D, et al. Application of clinical trial simulation to compare proof-of-concept study designs for drugs with a slow onset of effect; an example in Alzheimer's disease. Pharm Res 2006; 23(9): 2050-9.

Ludden TM, Beal SL, Sheiner LB. Comparison of the Akaike Information Criterion, the Schwarz criterion and the F test as guides to model selection. J Pharmacokinet Biop 1994;22(5):431-45.

Lunn DJ, Thomas A, Best N, et al. WinBUGS - a Bayesian modeling framework: concepts, structure, and extensibility. Stat Comput 2000;10:325-37.

Mohs RC, Rosen WG, Davis KL. The Alzheimer's disease assessment scale: an instrument for assessing treatment efficacy. Psychopharmacol Bull 1983;19(3):448-50.

Mould DR, Denman NG, Duffull S. Using disease progression models as a tool to detect drug effect. Clin Pharmacol Ther 2007; 82(1): 81-6.

Olanow CW, Rascol O, Hauser R, et al. A Double-Blind, Delayed-Start Trial of Rasagiline in Parkinson's Disease, New Engl J Med 2009;361(13):1268-78.

Olkin I, Tate RF. Multivariate correlation models with mixed discrete and continuous variables. Ann Math Stat 1961;32(2):448-65.

Putt ME, Ravina B. Randomized, placebo-controlled, parallel group versus crossover study designs for the study of dementia in Parkinson's disease. Control Clin Trials 2002; 23:111-26.

Qiu R, Rogers JA, Polhamus D, Romero K, Ito K, Corrigan B. Clinical trial simulation in Alzheimer's Disease: example applications of a modeling and simulation tool in drug Development. ASCPT, March 2012.

R Development Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2008.

Rogers, J, Polhamus, D, Ito, K, Romero, K, Qiu, R, Gillespie, Corrigan, B. Integration of Data For a Model-based Longitudinal Meta-analysis of Cognitive Changes in Mild and Moderate AD Patients. AAIC, July 2012a.

Rogers JA, Polhamus D, Gillespie WR, Ito K, Romero K, Qiu R, Stephenson D, Gastonguay MR, Corrigan B: Combining patient-level and summary-level data for Alzheimer's disease modeling and simulation: a beta regression meta-analysis. *J Pharmacokinet Pharmacodyn* 2012b, 39:479-498.

Rogers JA, Polhamus D, Ito K, Romero K, Qiu R, Gillespie B, Corrigan B. The value of evidence synthesis: model-based meta-analysis based on the CAMD database, the ADNI cohort and literature meta-data ASCPT, March 2012c.

Rogers, Jim, Polhamus, Dan, Lockwood, Peter, Brault, Yves, Desmet, Anne, Ito, Kaori, Romero, Klaus, Qiu, Ruolun, Gillespie, Bill, Corrigan, Brian, Gastonguay, Marc. Model- based Analysis to Support Strategic Decision Making: A Case Study from the Development of a 5HT6 Antagonist for the Treatment of Alzheimer's Disease. AAIC, July 2012d.

Romero K, Stephenson D, Hudson L, Rogers J, Polhamus D, Ito K, Qiu R, Corrigan B. Qualification of drug-disease trial models under the drug development tool (DDT) draft guidance: CAMD experience. ASPCT 2012a.

Romero K, Stephenson D, Hudson L, Rogers J, Polhamus D, Ito K, Qiu R, Corrigan B. The Coalition Against Major Diseases: Advancement of a Drug-Disease-Trial Model for Alzheimer's Disease Through a Regulatory Science Path. Arizona Alzheimer's Consortium, July 2012b.

Romero, Klaus, Stephenson, Diane, Rogers, James, Polhamus, Daniel, Ito, Kaori, Qiu, Ruolun, Corrigan, Brian. Advancement of a drug-disease-trial model for Alzheimer's Disease through a regulatory science path: the CAMD experience. AAIC, July 2012c.

Rosen WG, Mohs RC, Davis KL. A new rating scale for Alzheimer's disease. Am JPsychiat 1984; 141:1356-64.

Samtani MN, Farnum M, Lobanov V, et al. An improved model for disease progression in patients from the Alzheimer's Disease Neuroimaging Initiative. J Clin Pharmacol. 2012;52(5):629-44.

Schneider LS, Sano M. Current Alzheimer's disease clinical trials: methods and placebo outcomes. Alzheimers Dement 2009; 5:388-97.

Stukel TA. Generalized logistic models. J Am Stat Assoc 1988;83(42):426-31.

Tsoularis A, Wallace J. Analysis of logistic growth models. Math Biosci 2002;179(1):21-55.

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):120-31.

[i] All annexes mentioned under the Applicant's position refer to the documentation submitted with the request.