Featured Article

Design of a Bayesian adaptive phase 2 proof-of-concept trial for BAN2401, a putative disease-modifying monoclonal antibody for the treatment of Alzheimer's disease

Andrew Satlin[a],[*], Jinping Wang[a], Veronika Logovinsky[a], Scott Berry[b], Chad Swanson[a], Shobha Dhadda[a], Donald A. Berry[b]

[a]Neuroscience & General Medicine, Eisai Inc., Woodcliff Lake, NJ, USA
[b]Berry Consultants, LLC, Austin, TX, USA

Abstract

Introduction: Recent failures in phase 3 clinical trials in Alzheimer's disease (AD) suggest that novel approaches to drug development are urgently needed. Phase 3 risk can be mitigated by ensuring that clinical efficacy is established before initiating confirmatory trials, but traditional phase 2 trials in AD can be lengthy and costly.

Methods: We designed a Bayesian adaptive phase 2, proof-of-concept trial with a clinical endpoint to evaluate BAN2401, a monoclonal antibody targeting amyloid protofibrils. The study design used dose response and longitudinal modeling. Simulations were used to refine study design features to achieve optimal operating characteristics.

Results: The study design includes five active treatment arms plus placebo, a clinical outcome, 12-month primary endpoint, and a maximum sample size of 800. The average overall probability of success is ≥80% when at least one dose shows a treatment effect that would be considered clinically meaningful. Using frequent interim analyses, the randomization ratios are adapted based on the clinical endpoint, and the trial can be stopped for success or futility before full enrollment.

Discussion: Bayesian statistics can enhance the efficiency of analyzing the study data. The adaptive randomization generates more data on doses that appear to be more efficacious, which can improve dose selection for phase 3. The interim analyses permit stopping as soon as a predefined signal is detected, which can accelerate decision making. Both features can reduce the size and duration of the trial. This study design can mitigate some of the risks associated with advancing to phase 3 in the absence of data demonstrating clinical efficacy. Limitations to the approach are discussed.

© 2016 Eisai, Inc. Published by Elsevier Inc. on behalf of the Alzheimer's Association. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Keywords: Bayesian analysis; Adaptive trial; Alzheimer's disease; Interim analysis; Monoclonal antibody; Clinical trial simulation

1. Introduction

The last 16 putative disease-modifying agents for Alzheimer's disease (AD) failed to meet the primary efficacy objective in phase 3 trials (Table 1) [1–34].

Traditionally, confirmatory phase 3 trials follow successful phase 2 studies. The goals of phase 2 include establishing proof-of-concept using an endpoint that is appropriate for phase 3 or a suitable surrogate, demonstrating a dose-response that can be used to select the doses for phase 3, and assessing the magnitude of the treatment effect, so phase 3 studies can be appropriately powered.

This traditional strategy has not been followed in many recent AD projects for disease-modifying drugs [34]. Instead, small phase 2 studies using biomarker endpoints

Table 1
Summary of phase 2 and 3 trials of putative disease-modifying agents for the treatment of Alzheimer's Disease

Study agent	Phase	Subjects/arms	Duration	Outcome/comment	Reference
Atorvastatin	2	67/2	12 mo	Significant difference from placebo for ADAS-Cog seen at 6 months with trend toward significance for ADAS-Cog, CIBIC and NPI at 12 mo. No deterioration in cognitive and functional tests	[2]
	3	640/2	72 wk	No treatment benefit on ADAS-Cog or CGIC	[3]
Bapineuzumab	2	234/5	18 mo	No change on primary outcomes: ADAS-Cog11 and DAD. Positive trend for cognition in Apoε4-completers when all dose cohorts combined. Trend for reduction in CSF p-tau	[4]
	3	4000 in four studies 1. 1121/2 2. 1331/2 3. 835/3 4. 1099/2	78 wk	Negative overall on ADAS-Cog11 and DAD in first two studies. Project discontinued after first two studies reported	[5]
Indomethacin	2	51/2	12 mo	No treatment benefit on ADAS-Cog or other cognitive measures; results deemed inconclusive. Study was underpowered due to high dropout rate	[6]
	3	160 (estimated)/2	12 mo	Study completed. No results reported	NCT00432081
IVIG	2	24/2	6 mo	Statistically superior CGIC and numerically superior ADAS-Cog	[7]
	2	58/8	6 mo	No positive clinical outcomes	[8]
	3	390/3	18 mo	No treatment benefit on ADAS-Cog or ADCS-ADL	[9]
Latrepirdine/ dimebon	2	183/2	26 wk	Significant difference from placebo on ADAS-Cog. No biomarker analyses performed	[10]
	3	598/3	26 wk	No treatment benefit on ADAS-Cog or CIBIC+	[11]
Naproxen	2/3	351/3	12 mo	No treatment benefit on ADAS-Cog	[12]
	3	2528/3	5–7 y	No treatment benefit on cognition; possible detrimental effect. Study terminated early due to reported cardiovascular risk with celecoxib in another trial	[13]
Phenserine	2b	20/2	6 mo	Reduced CSF Aβ levels; significant effect on composite neuropsychological test	[14]
	3	384/3	6 mo	No significant treatment benefit on ADAS-Cog or CIBIC. Developmental program terminated	[15,16]
Rosiglitazone	2	30/2	6 mo	Decreased plasma Aβ in placebo group; improved delayed recall and selective attention in treatment group	[17]
	3	693/4	24 wk	No treatment benefit on ADAS-Cog or CIBIC+	[18]
Semagacestat	2	51/2	14 wk	Reduction in plasma Aβ40; no change in CSF Aβ. Suggestion of cognitive worsening	[19]
	3	2600 in two studies 1. 1537/3 2. 1111/2	21 mo	Deterioration on primary outcomes: ADAS-Cog11 and ADCS-ADL studies stopped at interim analysis	[20] NCT00762411
Simvastatin	2	44/2	26 wk	No overall change in CSF Aβ40/42 levels; post-hoc analysis showed decreased CSF Aβ40 in mild AD	[21]
	3	400/2	18 mo	No treatment benefit on ADAS-Cog	[21]
Solanezumab	2	52/5	12 wk	Dose-dependent increases of various Aβ species in plasma and CSF but no effects on the ADAS-Cog	[22]
	3	2000 in two studies 1. 1012/2 2. 1040/2	18 mo	Negative overall on primary outcomes: ADAS-Cog11 and ADCS-ADL. Preplanned analysis of combined study populations showed 34% reduction in rate of cognitive decline. Additional phase 3 program initiated based on this analysis	[23]
	3	Ongoing ~2100/2	18 mo	Primary endpoints: ADAS-Cog14; ADCS-iADL	NCT01900665
	2/3	Ongoing ~210/4	24 mo	Primary endpoint: changes in Aβ species	NCT01760005
	3	Ongoing_ ~1150/2	168 wk	Primary endpoint: ADCS-PACC	NCT02008357
Tarenflurbil	2	210/3	12 mo	Overall negative on primary endpoints: ADAS-Cog, ADCS-ADL, CDR-SB. Benefit on function in high-dose subjects with mild dementia at study entry	[24]
	3	1684/2	18 mo	No treatment benefit on ADAS-Cog or ADCS-ADL	[25]

(Continued)

Table 1
Summary of phase 2 and 3 trials of putative disease-modifying agents for the treatment of Alzheimer's Disease (*Continued*)

Study agent	Phase	Subjects/arms	Duration	Outcome/comment	Reference
Tramiprosate/	2	58/4	12 wk	Reduction in CSF Aβ42	[26]
Alzhemed	3	1052/3	18 mo	A trend toward improvement on ADAS-Cog; no change on CDR-sb	[27]
Valproate	2	n/a		No phase 2 performed. Hypothesized to have neuroprotective effects; further study implemented by ADCS	[28]
	3	313/2	24 mo	No clinical benefits of treatment; significant adverse events	[29]
VP4896 (leuprolide acetate, Memryte)	2	109 females/3 $\sim$90 males/3	48 wk	No treatment benefit on ADAS-Cog, ADCS-CGIC, or ADCS-ADL; subgroup of one study (women) showed maintenance of scores in latter 24 weeks	[30] NCT00076440
	3	555 (estimated)/ not provided	50 wk	Completed but data not reported	NCT00231946
Xaliproden	2	n/a		No phase 2 studies conducted in AD; proposed use in AD based on 5-HT1A antagonist mechanism	[16]
	3	2700 in two studies 1. 1455/2 2. 1306/2	18 mo	No treatment effect on ADAS-Cog or CDR; less hippocampal atrophy reported in subset of one study Development program discontinued	[16,31,32]

led directly to phase 3 trials (Table 1) [35,36]. These trials were generally underpowered to detect changes in clinical endpoints. Although biomarkers can indicate target engagement, this evidence alone is not sufficient to predict clinical effect and should not be used to advance a compound into phase 3 without evidence that the treatment's effect on the biomarker translates into a beneficial clinical effect. Although there are many possible reasons for the recent failed phase 3 trials of amyloid-based treatments, such as inclusion of patients at advanced stages of disease, one possible reason is the lack of clear evidence for an effect on a clinical endpoint in phase 2, leading to unsupported decisions to enter phase 3. To improve the success rate in phase 3 for a disease-modifying drug in AD, phase 2 trials should demonstrate a relevant treatment effect on a clinical endpoint.

AD trials using clinical endpoints will be larger and longer than those using biomarker endpoints for two reasons. First, clinical assessments of AD symptoms and progression have higher variability than biomarker assessments [37]. Second, disease-modifying AD therapies are generally tested earlier in the disease, when clinical progression is slower [38]. Thus, conducting traditional phase 2 clinical trials for disease-modifying therapies in AD is lengthy and costly. These considerations may have led companies to rely on data from smaller phase 2 studies using biomarker endpoints for decision making about entry into phase 3 and the design of phase 3 studies, with outcomes that have so far suggested the risk inherent in this approach.

Different trial designs are needed so that trials with clinical endpoints can be conducted more efficiently. An adaptive design can minimize the overall sample size and study duration by stopping recruitment early in response to strong signals of success or futility based on regular interim analyses (IAs) [39–41]. This approach can be enhanced by using the emerging data to adjust the randomization ratios to assign more subjects to doses that appear more efficacious, and fewer to doses that are less effective. This reduces the overall sample size and speeds recruitment by increasing the attractiveness of the trial to health authorities, IRBs, investigators, and subjects.

An important challenge of adaptive trials in AD is the difficulty of assessing the outcome measure early enough so modifications of the randomization can occur well before recruitment is complete [39,42]. In AD, where the clinical drug effect may not be seen for a long time, the theoretical advantages of adaptation have not been considered practical. This challenge can be met using a Bayesian design in which all available longitudinal data from the ongoing study are analyzed, with imputation of missing endpoint data based on the longitudinal model, so that response adaptive randomization and detection of signals predictive of 12-month success or futility can occur based on IAs before all subjects reach trial completion [42,43]. Early endpoints need not be surrogates for later endpoints to be used for imputation because this data-driven approach is not dependent on clinical hypotheses about the relationship of the endpoints. The longitudinal model is updated using data accruing in the trial, with correlations between observations at different time points informing the model. This approach results in a more efficient use of the data, leading to faster and more accurate decision making.

Based on these ideas, we designed a Bayesian adaptive phase 2, proof-of-concept trial with a clinical endpoint to evaluate BAN2401, a monoclonal antibody against amyloid protofibrils. This trial is currently underway (ClinicalTrials.gov identifier: NCT01767311).

2. Methods

2.1. Study objective

The primary objective of this study is to establish the effective dose 90% (ED$_{90}$) on the Alzheimer's Disease

Composite Clinical Score (ADCOMS) [44] at 12 months of treatment in subjects with early AD, defined as meeting NIA-AA criteria for mild cognitive impairment due to AD–intermediate likelihood or mild AD dementia. All subjects must have evidence for brain amyloid. These subjects are believed most likely to respond to an agent that neutralizes and clears toxic amyloid protofibrils in the brain. The ED_{90} for BAN2401 is defined as the simplest dose that achieves $\geq 90\%$ of the treatment effect achieved by the maximum effective dose, where the simplest dose is defined as the smallest dose with the lowest frequency of administration. ADCOMS is a novel instrument developed to improve the sensitivity of currently available cognitive and functional measures for subjects in the prodromal stage of AD. Based on the antibody pharmacokinetics, preclinical pharmacology, and safety results from phase I studies, we chose doses of 5 and 10 mg/kg monthly, and 2.5, 5, and 10 mg/kg biweekly, administered intravenously [45].

2.2. Fixed trial characteristics

The primary endpoint is change in ADCOMS from baseline to 52 weeks, which is considered long enough to detect clinically relevant changes. Follow-up continues through 18 months of treatment to detect effects on neuroimaging biomarkers, persistence of the clinical effect, or a pattern of longitudinal change suggestive of a disease-modifying effect. Efficacy assessments are performed at 13, 27, 39, and 52 weeks. We assumed a maximal recruitment rate of 32 subjects per month and a 20% dropout rate by 52 weeks.

We hypothesized a treatment effect equivalent to a 25% reduction in the rate of decline over 1 year. This effect was chosen to reflect a clinically significant difference (CSD) from placebo.

2.3. General approach to designing the adaptive trial

The prospectively defined adaptive features were response adaptive randomization and early termination of the study for futility or a strong signal for success. Both adaptations are based on the results of frequent IAs.

Adaptive randomization begins after a fixed allocation period. At each IA, the adaptive randomization probability for each of the five active doses changes based on the probability of each being the ED_{90}. The adaptive randomization probability for placebo mirrors the probability for the most likely ED_{90}.

Efficacy is tested against predefined stopping rules for futility or study success at each IA. The prespecified algorithm is run at each IA, the randomization probabilities are adapted, and the stopping rules are applied without any unblinding by study personnel.

Modeling and simulation were used to design this Bayesian adaptive trial. There are two major components to the modeling. The first is the modeling of the mean change from baseline to week 52 for each treatment arm. This model is informed by a second, longitudinal model that captures the correlations between early visits and the 52-week outcome. These two models work simultaneously to jointly estimate the primary endpoint effect size for each treatment arm at 52 weeks.

The dose-frequency response model and the longitudinal model were prospectively defined as part of the trial's design. They are stochastic models in that they have unknown parameters with probability distributions. These prior probability distributions were specified in advance and are updated as the trial proceeds and information accrues about each subject's outcome. This updating is Bayesian in nature.

Many iterations of the trial were simulated before finalizing the design. We assigned particular sets of numerical values to the parameters in the models and simulated an entire trial for many such sets to test a design's performance. Using the particular numerical values of the parameters, we generated each virtual subject's outcomes, including longitudinal outcomes, depending on the particular dose assigned to that subject. The computer algorithm that is the trial's design was not privy to the values of the parameters that generated the subject outcomes but only to the subject outcomes themselves, just as in the actual trial. The point of these simulations was to see how well a design was able to estimate the true values of the parameters. As these true values were known, we could compare them with the results of the simulations. Of course, in the actual trial, the true values of the parameters are not known.

2.4. Dose-frequency response model

A dose-frequency response model was constructed for the mean change from baseline to 52 weeks for each treatment arm (see Appendix A). The arms were modeled with a two-dimensional first-order normal dynamic linear model [46,47]. This model is a Gaussian random walk model over the dimension of dose and, separately, frequency of dose. The model allows a borrowing or smoothing of the 52-week effect size across neighboring doses and frequencies to provide a superior estimate of each treatment arm's effect. Prior distributions were set for the mean response for each dose and the error variance. The priors were set to be weak to allow the emerging data to shape the dose-frequency response model.

2.5. Longitudinal model

To allow the dose-frequency response model to be updated at each IA before 52-week data from all subjects are available, a linear regression longitudinal model was created for the correlation between the 6-week, 13-week, 27-week, and 39-week ADCOMS values and the 52-week value (see Appendix A). The model was built to learn from the accruing information during the course of the trial on the observed

correlation between the earlier values and the 52-week primary value using Bayesian imputation within Markov chain Monte Carlo, to impute the 52-week data and update the dose-frequency response model at each IA. The joint modeling is done using imputations within the longitudinal model to allow joint estimation of the dose-frequency response [47,48]. The prior distributions were selected from historical data (mean slope of 0.80) as empirically based starting values but were kept weak to allow the new study data to shape the posterior distribution.

2.6. Simulations used in the design of the trial

Simulations are a critical part of designing a Bayesian adaptive trial: they allow a full exploration of the design and the ability to calculate its operating characteristics, as these are not available analytically as in a traditional, fixed-design trial. Key operating characteristics of interest were the type I error and the probabilities of futility, early success, overall success, selecting each treatment arm for phase 3, and success in a phase 3 confirmatory trial. The simulated virtual trials allowed us to calculate the operating characteristics of the trial for a wide range of possible dose-frequency response scenarios. Simulations were used to iterate aspects of the design, including maximum sample size, cut-offs for stopping rules, frequency of IAs, and response adaptive randomization, to achieve optimized operating characteristics. If, as new design features were simulated, these did not result in an optimal balancing of operating characteristics (e.g., between the probabilities of stopping for futility or early success in different scenarios), the iterative process was resumed to correct them [40,42]. This iterative design process and exploration of operating characteristics were used to create a final, completely prospectively defined protocol [42,49,50].

Simulations were performed using scenarios that included a strong null scenario, in which no dose had any effect, and a large number of dose-frequency response scenarios reflecting possible outcomes for relative efficacy among the treatment arms and for when treatment effects might emerge (see Appendix B). The scenarios in which some doses are effective and some are not included five exposure-response relationships: (1) linear dose-response throughout the dose range; (2) linear dose-response that is only seen above a threshold dose of 10 or 20 mg/kg/month; (3) flat dose-response (all doses work equally well); (4) inverted U-shaped dose-response (intermediate doses have the best effect, and a lesser effect is achieved with doses that are either higher or lower); and (5) dose response based on a need to maintain a minimum average drug concentration (biweekly dosing works better than monthly dosing). Within each of these dose-frequency response pattern scenarios, we included several subscenarios with treatment effect magnitudes ranging from very weak to very strong reductions in the rate of clinical decline.

Based on assumptions about the fraction of the 52-week treatment effect that would be observed at each visit, we chose three different longitudinal patterns of drug response: (1) linear (assumes steady growth in treatment effect through 52 weeks; essentially, the pattern that should be seen if the drug is disease-modifying); (2) symptomatic (most of the effect occurs in the first 12 weeks, then remains the same); and (3) late-onset (which assumes that a certain amount of amyloid-based toxicity must be cleared before a clinical effect is seen). Each dose-response scenario was simulated under the assumption of each of these patterns of response.

An iterative simulation process was used to choose the futility and success boundaries. We defined futility as a probability $<X\%$ that any dose is better than placebo by the CSD at an IA, and early success as a probability of $\geq Y\%$ that any dose is better than placebo by the CSD. We simulated the trial using a range of 2.5%–15% probability of no effect for futility and 85%–99% probability that one dose was efficacious for early success. Based on the simulations, we chose boundaries that resulted in the most useful balance between improving trial efficiency and minimizing the risk of making incorrect decisions about stopping the trial early (see Fig. 1 and 2 for detailed examples; note that the type I error rate of 16% in Fig. 2 is the result at the beginning of the design process. After simulations led to improvements in the initial design, the final type I error rate was about 10%).

Because enrollment in the study can be stopped once a signal is reached at any IA, the actual sample size is not prespecified in a Bayesian adaptive trial. Instead, a maximum sample size is chosen. Simulations were used to select the maximum sample size of the trial. Sample sizes of 500–900 were simulated. Under the stopping rules that provided the most assurance of trial success, the best balance of power and time to obtain a signal of success was seen with a maximum sample size of 800. This maximum is also the anticipated size of a traditional fixed-design trial, so that if there is no interim signal for stopping, the trial will still have power to achieve the objective once all subjects have completed treatment.

Recruitment rate is an important factor in designing the trial because it affects the amount of data that will be available at each interim analysis (i.e., because the IAs are conducted after every 50 subjects are randomized, the recruitment rate determines the treatment duration for subjects previously randomized, and therefore, the total amount of data available for analysis). Therefore, the improved efficiency due to the Bayesian analyses is tempered as the recruitment rate increases, such that the advantage of a faster recruitment rate in accelerating the conduct of the trial is somewhat lessened. Recruitment rate can therefore be adjusted to maximize trial efficiency, and simulations can be performed to determine the optimal recruitment rate. However, in practice, recruitment is always challenging in dementia trials, especially ones seeking subjects with prodromal AD. Therefore, as noted above, we chose to fix the recruitment rate at the maximum we expected to achieve.

A. Satlin et al. / Alzheimer's & Dementia: Translational Research & Clinical Interventions 2 (2016) 1-12
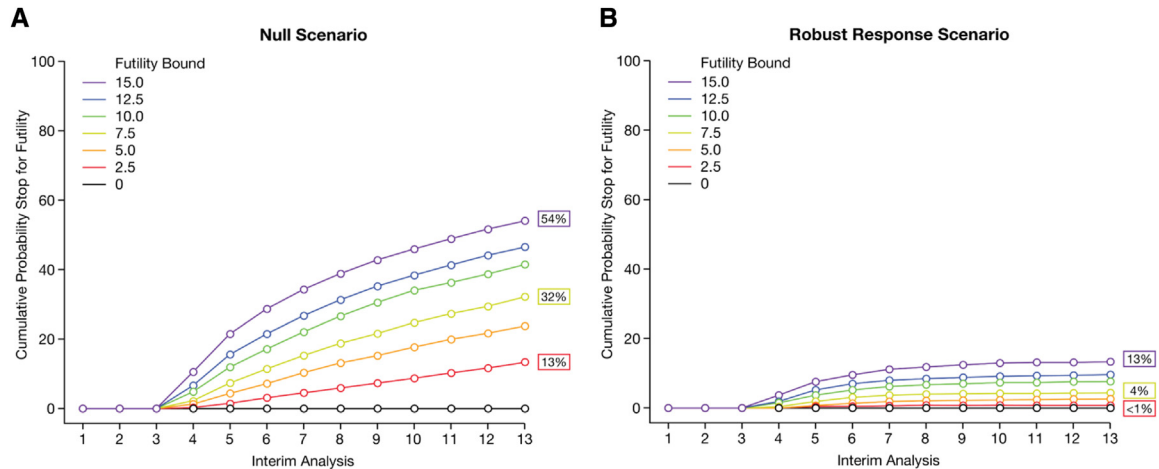

Fig. 1. Simulating futility boundaries in multiple dose and effect scenarios. Futility boundaries ranging from 2.5% to 15% were simulated for each scenario. The results for two scenarios are shown: (A) null scenario and (B) a dose-response scenario in which one dose has a robust effect. Robust indicates a dose-response in which the percentage reduction in decline relative to placebo for the 2.5-mg bimonthly, 5-mg bimonthly, 10-mg bimonthly, 5-mg monthly, and 10-mg monthly doses are 17%, 33%, 50%, 17%, and 33%, respectively. Null scenario simulations showed that with a boundary of 15%, the cumulative probability of declaring futility at the 13th IA would be 54% (A). However, using the same boundary, futility could be declared 13% of the time in a scenario where a single dose had a robust response (B), which was considered too risky. A boundary of 2.5% would reduce the probability of declaring futility to nearly zero in the event of a robust response for one dose (B) but would only permit stopping for futility in the null scenario 13% of the time (A). A boundary of 7.5% would permit stopping for futility 32% of the time in the null scenario (A), while only incorrectly declaring futility 4% of the time when one dose actually had a robust effect (B). Based on these simulation results, a futility boundary of 7.5% was chosen as providing the most acceptable balance of trial efficiency and risk. For the first three IAs (i.e., at 196, 250, and 300 subjects randomized), a more conservative futility boundary of 5% was chosen to further reduce the possibility of inappropriately stopping early for futility at very early time points in the trial, when the decision would be based on more limited data.

3. Results

The final study design, using five active treatment arms and placebo, ADCOMS outcome measure, 12-month primary endpoint, and a maximum sample size of 800, has an average overall probability of success of ≥80% when at least one dose shows a treatment effect that would be considered clinically meaningful.

In this phase 2 design, a type I error would be a decision to move to a phase 3 trial when there is no cognitive benefit of the drug. Simulations of the possible outcomes indicated that the type I error rate (one sided) is ∼10%, which we considered acceptable for a proof-of-concept trial. Note that, in a frequentist approach, the P value is calculated using a test statistic based on the distribution center at 0 under the null hypothesis, which is naturally two sided, and a one-sided P value of 10% is equivalent to a two-sided P value of 20%. However, the type I error rate in this Bayesian adaptive trial is not based on any statistical test of significance, but simply on the false-positive rate of success under the null hypothesis through simulations. Therefore, in this Bayesian trial, the simulated type I error rate is the experimental wise error rate, which in this design is 10%.

The mean actual sample size averaged across all scenarios is expected to be 626. A decision on success or failure, and therefore the opportunity to advance to phase 3 or discontinue the program would occur, on average, 17 months earlier than for a traditional trial with 800

subjects. This gain in efficiency is due not only to stopping enrollment before the maximum sample size is reached, but also to being able to declare success before all subjects have completed the full 12 months of treatment.

The protocol design is presented as a figure in Appendix C. An example of a simulated virtual trial is presented in Fig. 3. This example illustrates the response adaptive randomization and emergence of potential signals of early success or futility.

IAs occur when 196 subjects are enrolled, then 250, and then every 50 subjects up to the maximum sample size of 800. After 800 subjects have been randomized, if neither boundary has been crossed, IAs occur every 3 months. The 52-week ADCOMS value is the driver of all decisions and/or adaptations at each IA. For subjects that do not yet have a 52-week observation, the 52-week observation is calculated using multiple imputations from the longitudinal model based on the most recent observation. At each IA, a Bayesian analysis is used to calculate the posterior distribution of the parameters in the dose-frequency response model. Based on these distributions, 10,000 dose-response curves are sampled. The maximum effective dose and the ED_{90} for each of these curves are identified. The proportion of simulated dose–frequency response curves where a given arm is the maximum effective dose is the posterior probability that that dose is the maximum effective dose. Similarly, the
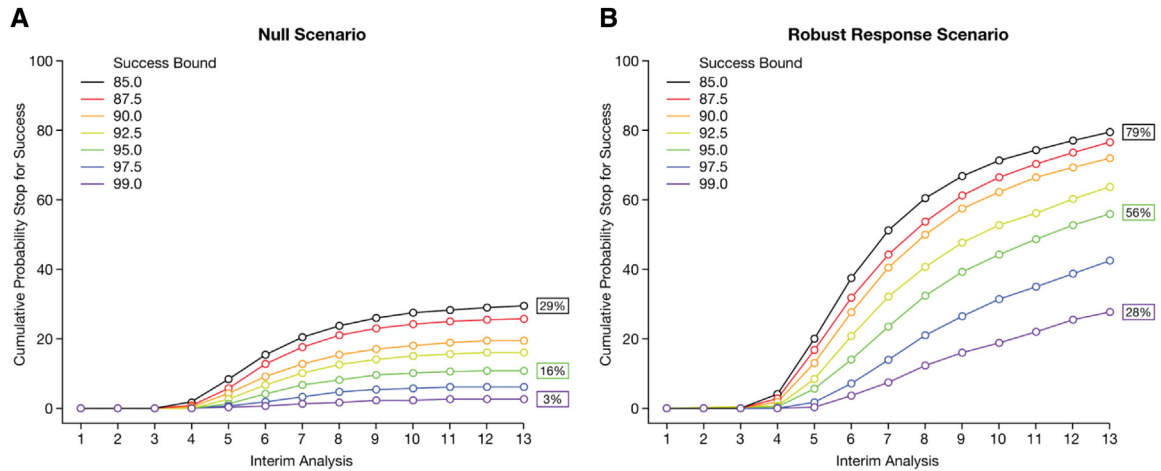
Fig. 2. Simulating early success boundaries in multiple dose and effect scenarios. Early success boundaries ranging from 85% to 99% were simulated for each scenario. The results for two scenarios are shown: (A) null scenario and (B) dose-response scenario in which one dose has a robust effect. Robust indicates a dose-response in which the percentage reduction in decline relative to placebo for the 2.5-mg bimonthly, 5-mg bimonthly, 10-mg bimonthly, 5-mg monthly, and 10-mg monthly doses are 17%, 33%, 50%, 17%, and 33%, respectively. With a success boundary of 85%, the cumulative probability of declaring early success at the 13th IA was 79% if one dose had a robust effect (B), but this boundary would result in success being declared 29% of the time in the null scenario (A), which was considered too risky. A success boundary of 99% would reduce the incorrect declaration of success in the null scenario to 3% (A) but would only permit stopping early for success when one dose had a robust effect 28% of the time (B). A success boundary of 95% would declare early success 56% of the time when one dose had a robust effect (B), while incorrectly declaring success 16% of the time in the null scenario (A). Based on these simulation results, an early success boundary of 95% was chosen as providing the most acceptable balance of trial efficiency and risk.

proportion of simulated dose–frequency response curves where a given arm is the ED_{90} is the posterior probability that that dose is the ED_{90}. These probabilities are used for an algorithm-determined adjustment of the randomization allocation probabilities, implemented in a computerized system that automatically assigns treatments to subsequent subjects without unblinding any study or sponsor personnel or the study investigators.

Similarly, for each active treatment arm, the posterior distribution of the mean change from baseline to 52 weeks is compared with placebo to calculate the posterior probability of being superior to placebo and of being superior to placebo by at least the CSD. These probabilities are used to determine whether the trial has met the predefined criteria for early stopping.

Interim monitoring for futility begins at the first IA and is based on the dose identified as the most likely ED_{90}. At the first three IAs, if there is a <5% posterior probability that the most likely ED_{90} is superior to placebo by the CSD, the trial will stop early for futility. Beginning at the 350-subject IA, and continuing to completion of the trial, the futility criterion is increased to 7.5%. Thus, it becomes easier to stop for futility once ≥350 subjects have been enrolled. If a signal for futility is found, the sponsor is notified and the trial is terminated.

Interim monitoring for early success occurs at each IA beginning when 350 subjects have been enrolled. At this point, if enrollment were to stop for early success, enough subjects would be available to complete the trial so that the full dose response could be modeled. If there is a >95% posterior probability that the most likely ED_{90} is

better than placebo by the CSD, then early success is declared. Enrollment is stopped, but all randomized subjects continue for the full 18-month duration of the study.

If the trial is not stopped early for futility or success, then trial success is evaluated at the completion of the trial, when both accrual and follow-up for the primary endpoint are complete. At that time, if there is a >80% probability that the most likely ED_{90} is better than placebo by the CSD, the trial will be considered a success. If the trial is stopped early for success, then a final analysis using the full 12-month data from all subjects serves as a sensitivity analysis to confirm the result. Stopping early for success means that there is a 95% probability that the ED_{90} is better than placebo by the CSD. Therefore, if this occurs, there is a very high likelihood of achieving an 80% probability that the ED_{90} is better than placebo by the CSD at the final analysis. However, if the final sensitivity analysis does not confirm the result, additional analyses of the data would be performed to explain the discrepancy.

Unlike in a traditional trial, achieving study success is not the same as meeting the primary objective. Although the primary objective in this study is met by identifying the ED_{90}, study success requires that this dose can achieve a clinically meaningful treatment effect in a phase 3 study. In the phase 2 trial, success is defined as a drug effect that exceeds the placebo rate by ≥25%, rather than only being superior to placebo. We chose these criteria to ensure that any early signal of success would likely indicate a robust treatment effect. Thus, using the Bayesian adaptive design in phase 2, the risks of failure in phase 3 are mitigated.
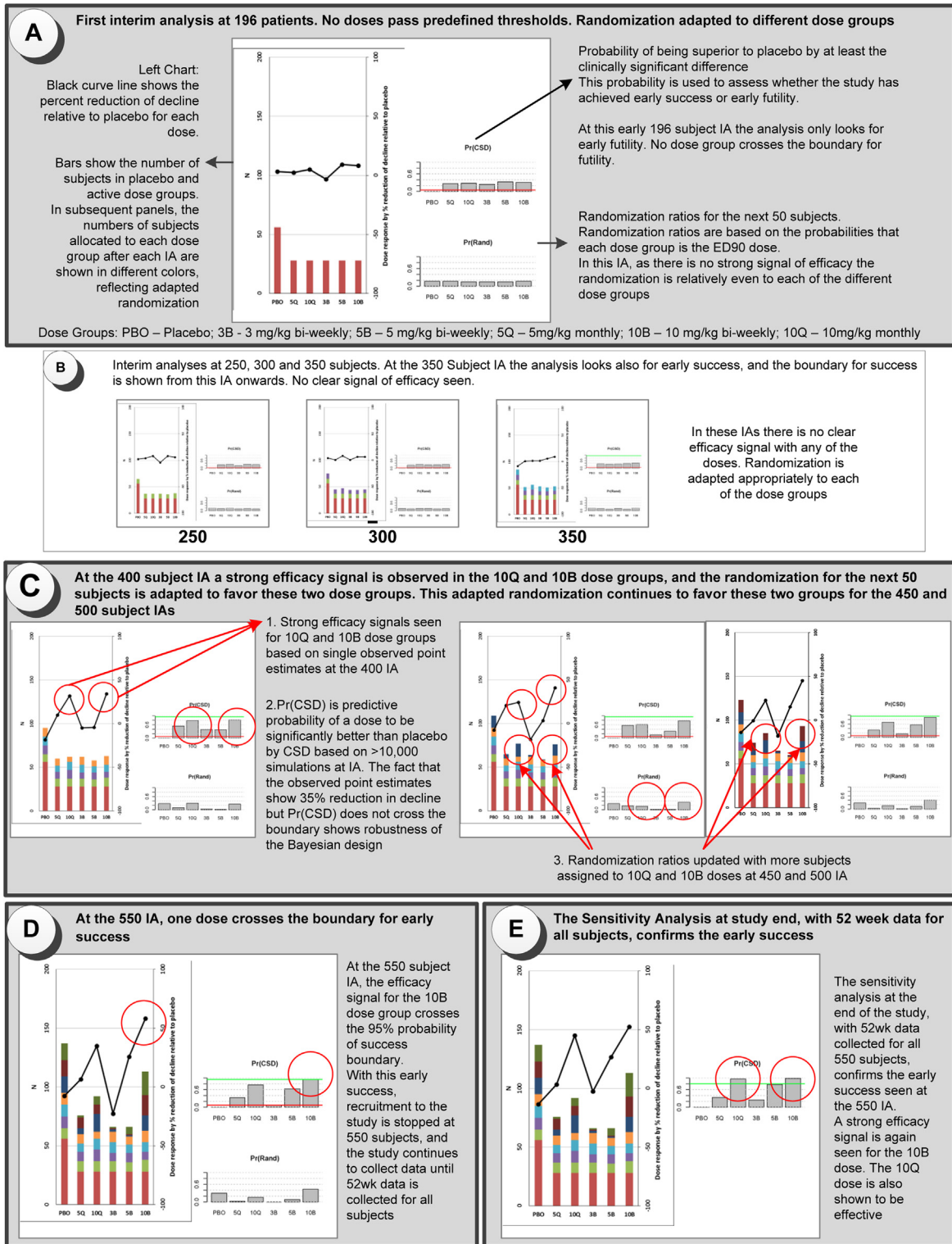
Fig. 3. One simulation of a robust dose-response scenario, in which the trial is stopped early for success at 550 subjects. Robust dose-response scenario used is the linear dose-response scenario with the largest response having ≥50% reduction in decline relative to placebo. During the study, IAs are carried out after 196 subjects are randomized (A), after 250 subjects, and every 50 subjects thereafter (B) until 800 subjects are recruited, after which IAs are carried out every 3 months, until 52-week data are available for all subjects. Based on the results of each IA, the study can be stopped early for success or futility. If the study continues, the randomization ratios are adapted based on the effects of each dose on the ADCOMS (C). If the study achieves early success, recruitment is stopped (D), and the study continues with enrolled subjects until 52-week data are collected for all subjects (E). Robust indicates a dose-response in which the percentage reduction in decline relative to placebo for the 2.5-mg bimonthly, 5-mg bimonthly, 10-mg bimonthly, 5-mg monthly, and 10-mg monthly doses are 17%, 33%, 50%, 17%, and 33%, respectively.

4. Discussion

Many recent failures in phase 3 studies of AD drugs can be attributed to lack of robust phase 2 clinical results. Owing to the long-term endpoint, large variability in disease progression, typically high dropout rates and limited size of the potential treatment effect due to the nature of the drug (disease modification), the outcomes from shorter and smaller studies are likely to be misleadingly large or small. However, traditional phase 2 studies have been avoided in recent AD drug development because of concerns that they are too large and costly. This risk is a major challenge in developing disease-modifying drugs in AD. We believe that the study described in this article permits the determination of clinical efficacy for a disease-modifying treatment for AD in a trial of feasible size and duration by using a Bayesian adaptive design. The results can be used for go/no-go decisions on initiating phase 3 and to design phase 3 trials with a greater likelihood of success.

This trial design has numerous advantages over a traditional phase 2 proof-of-concept, dose-finding study and over other statistical approaches. Each Bayesian IA uses emerging data to update the joint posterior distribution of the longitudinal model parameters, which improves the predictive power of IAs [39,41]. In this way, the regular IAs are part of a learning process that improves the trial's efficiency in arriving at a stopping decision. This approach compensates for the absence of good predictive biomarkers or a priori evidence for the clinical effect of the study drug. A group sequential trial design permits IAs for early stopping but does not incorporate longitudinal data, a dose-response, or response adaptive randomization [40,51,52].

The study can stop early, either for futility or success, thereby avoiding unnecessary exposure of patients to an ineffective drug or accelerating the initiation of confirmatory trials that can speed the availability of an effective drug to patients who need it. This efficiency gain is due to the likelihood that enrollment will complete before the 800 subject maximum and due to the efficacy signal being detected before all enrolled subjects have completed the full 12 months of treatment. If the drug is ineffective, the time and cost savings can be diverted to more promising therapies.

There are a number of limitations to the use of the Bayesian adaptive approach for clinical trials. All Bayesian analyses are dependent on the choice of prior distributions, which can be erroneous. In this trial, the prior distributions for the dose-frequency response model and the longitudinal model were kept weak so that the posterior distributions would be mostly shaped by the emerging data. Our estimates of the operating characteristics were based on simulations of a range of possible dose-frequency response scenarios, but these may not have covered all possible scenarios, which could have affected our estimate of power if the true dose-response scenario was to be very different from what was simulated. Underestimation of the dropout rate could also result in underestimation of type I error rate. In addition, we could not be certain that each of the simulated scenarios was equally likely, so it is possible that our estimate for the average expected sample size, which was not weighted, might not be accurate. For these reasons, our estimates for the potential efficiency gains that could be achieved with this study design should be viewed with some caution. In addition, the conclusions of the trial could be impacted by the possibility that effect sizes observed at early success might be larger than the true effect size, resulting in a positive bias.

Designing a Bayesian adaptive trial requires extensive planning [39]. The study team must consider numerous scenarios for possible outcomes and must expend time and resources on a lengthy iterative simulation process. It should be noted, however, that all trials, whether adaptive or fixed, should be simulated to ensure the effectiveness of the design to achieve the study aims; this approach can minimize the risk of inadequate phase 2 trials and costly phase 3 failures. Conducting an adaptive trial requires a rapid flow of data from sites to the database so IAs can be performed in real time to adjust the randomization ratios on schedule and provide the study team with early stopping decisions if a boundary is met. Drug use and availability must be tracked more closely than for a traditional trial because of the adjustments in randomization ratios. Drug supplies may need to be larger than for a traditional trial because of these adjustments [39].

Regulators may be less familiar with these complex study designs [39], although the Food and Drug Administration has issued draft guidance [53]. Prior agreement with regulators on the study design is helpful. Questions about the maintenance of data integrity using this trial design are to be expected [40]. In the BAN2401 trial, once the protocol was finalized, the sponsor (Eisai) has had no involvement with the actual IAs. An independent monitoring committee and statistical group are involved with the implementation of the IAs, whereas Eisai and the investigators remain blinded to study treatment and changes in randomization. These procedures mitigate the risks of any compromise to study data integrity and maintain the type I error [40]. A properly conducted Bayesian adaptive study that achieves robust positive results on the primary endpoint could be discussed with regulators for possible use in supporting drug registration.

Two projects are underway to design standing adaptive proof-of-concept trials of therapies for prevention of AD (Innovative Medicines Initiative–European Prevention of Alzheimer's Dementia [IMI-EPAD] project; GAP [Global Alzheimer's Platform]) [54,55]. These efforts are led by consortia with access to many potential drug candidates. The design of the study described in this article could provide a framework for the clinical trial design aspects of these consortia projects.

Acknowledgments

Ltd, Brighton, UK. The study described in the article is funded by the study sponsor, Eisai, Inc., Woodcliff Lake, NJ, United States.

Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.trci.2016.01.001.

RESEARCH IN CONTEXT

1. Systematic review: A PubMed search was done to identify clinical trials for putative disease-modifying drugs for Alzheimer's disease (AD). Trials were reviewed for design, results, and the overall compound clinical trial development program.

2. Interpretation: The study design described in this article represents the first use of a Bayesian adaptive clinical trial design for determining proof-of-concept for a disease-modifying drug in AD. This design is feasible, although the primary endpoint comes after a year or more of treatment, and it can mitigate the risks of starting large and costly phase 3 clinical trials with limited data on clinical endpoints.

3. Future directions: The trial design described here should be considered for future phase 2 studies in AD or in other indications where the data needed to design trials with a fixed design are not available and initiation of large and costly phase 3 studies entails high risk.

References

[1] Cummings JL, Morstorf T, Zhong K. Alzheimer's disease drug-development pipeline: few candidates, frequent failures. Alzheimers Res Ther 2015;6:37–43.

[2] Sparks DL, Sabbagh MN, Connor DJ, Lopez J, Launer LJ, Browne P, et al. Atorvastatin for the treatment of mild to moderate Alzheimer disease: preliminary results. Arch Neurol 2005;62:753–7.

[3] Feldman HH, Doody RS, Kivipelto M, Sparks DL, Waters DD, Jones RW, et al. LEADe Investigators. Randomized controlled trial of atorvastatin in mild to moderate Alzheimer disease: LEADe. Neurology 2010;74:956–64.

[4] Salloway S, Sperling R, Gilman S, Fox NC, Blennow K, Raskind M, et al. Bapineuzumab 201 Clinical Trial Investigators. A phase 2 multiple ascending dose trial of bapineuzumab in mild to moderate Alzheimer disease. Neurology 2009;73:2061–70.

[5] Salloway S, Sperling R, Fox NC, Blennow K, Klunk W, Raskind M, et al. Bapineuzumab 301 and 302 Clinical Trial Investigators. Two phase 3 trials of bapineuzumab in mild-to-moderate Alzheimer's disease. N Engl J Med 2014;370:322–33.

[6] de Jong D, Jansen R, Hoefnagels W, Jellesma-Eggenkamp M, Verbeek M, Borm G, et al. No effect of one-year treatment with indomethacin on Alzheimer's disease progression: a randomized controlled trial. PLoS One 2008;3:e1475.

[7] Relkin N. Intravenous immunoglobulin for Alzheimer's disease. Clin Exp Immunol 2014;178(Suppl 1):27–9.

[8] Dodel R, Rominger A, Bartenstein P, Barkhof F, Blennow K, Förster S, et al. Intravenous immunoglobulin for treatment of mild-to-moderate Alzheimer's disease: a Phase 2, randomised, double-blind, placebo-controlled, dose finding trial. Lancet Neurol 2013;12:233–43.

[9] Relkin N. Clinical trials of intravenous immunoglobulin for Alzheimer's disease. J Clin Immunol 2014;34(Suppl 1):S74–9.

[10] Doody RSI, Gavrilova SI, Sano M, Thomas RG, Aisen PS, Bachurin SO, et al. Effect of dimebon on cognition, activities of daily living, behaviour, and global function in patients with mild-to-moderate Alzheimer's disease: a randomised, double-blind, placebo-controlled study. Lancet 2008;372:207–15.

[11] Bezprozvanny I. The rise and fall of Dimebon. Drug News Perspect 2010;23:518–23.

[12] Aisen PS, Schafer KA, Grundman M, Pfeiffer E, Sano M, Davis KL, et al. Alzheimer's Disease Cooperative Study. Effects of rofecoxib or naproxen vs placebo on Alzheimer disease progression: a randomized controlled trial. JAMA 2003;289:2819–26.

[13] ADAPT Research Group Martin BK, Szekely C, Brandt J, Piantadosi S, Breitner JC, Craft S, et al. Cognitive function over time in the Alzheimer's Disease Anti-inflammatory Prevention Trial (ADAPT): results of a randomized, controlled trial of naproxen and celecoxib. Arch Neurol 2008;65:896–905.

[14] Kadir A, Andreasen N, Almkvist O, Wall A, Forsberg A, Engler H, et al. Effect of phenserine treatment on brain functional activity and amyloid in Alzheimer's disease. Ann Neurol 2008;63:621–31.

[15] Axonyx. Axonyx announces that phenserine did not achieve significant efficacy in phase 3 Alzheimer's disease trial. Available at: http://www.businesswire.com/news/home/20050206005019/en/Axonyx-Announces-Phenserine-Achieve-Significant-Efficacy-Phase#.VLADX_R0xZQ. Accessed January 9, 2015.

[16] Sabbagh MN. Drug development for Alzheimer's disease: where are we now and where are we headed? Am J Geriatr Pharmacother 2009;7:167–85.

[17] Watson GS, Cholerton BA, Reger MA, Baker LD, Plymate SR, Asthana S, et al. Preserved cognition in patients with early Alzheimer disease and amnestic mild cognitive impairment during treatment with rosiglitazone: a preliminary study. Am J Geriatr Psychiatry 2005;13:950–8.

[18] Gold M, Alderton C, Zvartau-Hind M, Egginton S, Saunders AM, Irizarry M, et al. Rosiglitazone monotherapy in mild-to-moderate Alzheimer's disease: results from a randomized, double-blind, placebo-controlled phase 3 study. Dement Geriatr Cogn Disord 2010;30:131–46.

[19] Fleisher AS, Raman R, Siemers ER, Becerra L, Clark CM, Dean RA, et al. Phase 2 safety trial targeting amyloid beta production with a gamma-secretase inhibitor in Alzheimer disease. Arch Neurol 2008;65:1031–8.

[20] Doody RS, Raman R, Farlow M, Iwatsubo T, Vellas B, Joffe S, et al, Alzheimer's Disease Cooperative Study Steering Committee, Semagacestat Study Group. A phase 3 trial of semagacestat for treatment of Alzheimer's disease. N Engl J Med 2013;369:341–50.

[21] Sano M, Bell KL, Galasko D, Galvin JE, Thomas RG, van Dyck CH, et al. A randomized, double-blind, placebo-controlled trial of simvastatin to treat Alzheimer disease. Neurology 2011;77:556–63.

[22] Farlow M, Arnold SE, van Dyck CH, Aisen PS, Snider BJ, Porsteinsson AP, et al. Safety and biomarker effects of solanezumab in patients with Alzheimer's disease. Alzheimers Dement 2012;8:261–71.

[23] Doody RS, Thomas RG, Farlow M, Iwatsubo T, Vellas B, Joffe S, et al. Phase 3 trials of solanezumab for mild-to-moderate Alzheimer's disease. N Engl J Med 2014;370:311–21.

[24] Wilcock GK, Black SW, Hendrix SB, Zavitz KH, Swabb EA, Laughlin MA, Tarenflurbil Phase 2 Study investigators. Efficacy and safety of tarenflurbil in mild to moderate Alzheimer's disease: a randomised phase 2 trial. Lancet Neurol 2008;7:483–93. Errata in Lancet Neurol 2008;7:575; Lancet Neurol 2011;10:297.

[25] Green RC, Schneider LS, Amato DA, Beelen AP, Wilcock G, Swabb EA, et al. Tarenflurbil Phase 3 Study Group. Effect of tarenflurbil on cognitive decline and activities of daily living in patients with mild Alzheimer disease: a randomized controlled trial. JAMA 2009; 302:2557–64.

[26] Aisen PS, Gauthier S, Vellas B, Briand R, Saumier D, Laurin J, et al. Alzhemed: a potential treatment for Alzheimer's disease. Curr Alzheimer Res 2007;4:473–8.

[27] Aisen PS, Gauthier S, Ferris SH, Saumier D, Haine D, Garceau D, et al. Tramiprosate in mild-to-moderate Alzheimer's disease – a randomized, double-blind, placebo-controlled, multi-centre study (the Alphase Study). Arch Med Sci 2011;7:102–11.

[28] Tariot PN, Loy R, Ryan JM, Porsteinsson A, Ismail S. Mood stabilizers in Alzheimer's disease: symptomatic and neuroprotective rationales. Adv Drug Deliv Rev 2002;54:1567–77.

[29] Tariot PN, Schneider LS, Cummings J, Thomas RG, Raman R, Jakimovich LJ, et al. Alzheimer's Disease Cooperative Study Group. Chronic divalproex sodium to attenuate agitation and clinical progression of Alzheimer disease. Arch Gen Psychiatry 2011; 68:853–61.

[30] Bowen RL, Perry G, Xiong C, Smith MA, Atwood CS. A clinical study of lupron depot in the treatment of women with Alzheimer's disease: preservation of cognitive function in patients taking an acetylcholinesterase inhibitor and treated with high dose lupron over 48 weeks. J Alzheimers Dis 2015;44:549–60.

[31] APM Health Europe. Sanofi's Alzheimer's drug xaliproden fails in phase 3. 2007. Available at: http://www.apmhealtheurope.com/story.php?mots=UPDATE&searchScope=1&searchType=0&numero=L8198. Accessed December 11, 2015.

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

[33] Salomone S, Caraci F, Leggio GM, Fedotova J, Drago F. New pharmacological strategies for treatment of Alzheimer's disease: focus on disease modifying drugs. Br J Clin Pharmacol 2012;73:504–17.

[34] Schneider LS, Mangialasche F, Andreasen N, Feldman H, Giacobini E, Jones R, et al. Clinical trials and late-stage drug development for Alzheimer's disease: an appraisal from 1984 to 2014. J Intern Med 2014; 275:251–83.

[35] Greenberg BD, Carrillo MC, Ryan JM, Gold M, Gallagher K, Grundman M, et al. Improving Alzheimer's disease phase 2 clinical trials. Alzheimers Dement 2013;9:39–49.

[36] Gray JA, Fleet D, Winblad B. The need for thorough phase 2 studies in medicines development for Alzheimer's disease. Alzheimers Res Ther 2015;7:67–9.

[37] Riepe MW, Wilkinson D, Förstl H, Brieden A. Additive scales in degenerative disease – calculation of effect sizes and clinical judgment. BMC Med Res Methodol 2011;11:169.

[38] Albert MS, DeKosky ST, Dickson D, Dubois B, Feldman HH, Fox NC, et al. The diagnosis of mild cognitive impairment due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimers Dement 2011; 7:270–9.

[39] Gallo P, Chuang-Stein C, Dragalin V, Gaydos B, Krams M, Pinheiro J, PhRMA Working Group. Adaptive designs in clinical drug development – an Executive Summary of the PhRMA Working Group. J Biopharm Stat 2006;16:275–83. discussion 285–91, 293–8, 311–12.

[40] Gallo P, Anderson K, Chuang-Stein C, Dragalin V, Gaydos B, Krams M, et al. Viewpoints on the FDA draft adaptive designs guidance from the PhRMA working group. J Biopharm Stat 2010;20:1115–24.

[41] Yin G, Chen N, Lee JJ. Phase 2 trial design with Bayesian adaptive randomization and predictive probability. J R Stat Soc Ser C Appl Stat 2012;61.

[42] Berry SM, Spinelli W, Littman GS, Liang JZ, Fardipour P, Berry DA, et al. A Bayesian dose-finding trial with adaptive dose expansion to flexibly assess efficacy and safety of an investigational drug. Clin Trials 2010;7:121–35.

[43] Berry DA. Adaptive clinical trials in oncology. Nat Rev Clin Oncol 2011;9:199–207.

[44] Hendrix S, Logovinsky V, Perdomo C, Wang J, Satlin A. Optimizing responsiveness to decline in early AD. Presented at AAIC, July 14–19, 2012; Vancouver, Canada. Abstract #P4-305.

[45] Lai R, Logovinsky V, Kaplow JM, Gu K, Yu Y, Möller C, et al. A first-in-human study of BAN2401, a novel monoclonal antibody against amyloid-β protofibrils. Presented at AAIC 2013; Boston, MA.

[46] West M, Harrison P. Bayesian Forecasting and Dynamic Models. 2nd ed. New York: Springer; 1997. p. 97–142. Chapter.

[47] Berry SM, Carlin BP, Lee JJ, Muller P. Bayesian Adaptive Methods for Clinical Trials. Boca Raton: CRC Press; 2011. p. 193–247. Chapter 5.

[48] Padmanabhan SK, Berry SM, Dragalin V, Krams M. A Bayesian dose-finding design adapting to efficacy and tolerability response. J Biopharm Stat 2012;22:276–93.

[49] Berry DA. Bayesian clinical trials. Nat Rev Drug Discov 2006;5:27–36.

[50] Berry DA. Interim analysis in clinical trials: the role of the likelihood principle. Am Stat 1987;41:117–22.

[51] Mehta CR, Patel NR. Adaptive, group sequential and decision theoretic approaches to sample size determination. Stat Med 2006;25:3250–69.

[52] Chuang-Stein C. Sample size and the probability of a successful trial. Pharm Stat 2006;5:305–9.

[53] Food and Drug Administration. Guidance for Industry: Adaptive design clinical trials for drugs and biologics. Available at: http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM201790.pdf. Accessed January 19, 2015.

[54] Ritchie CW, Lovestone S, Molinuevo JL, Diaz C, Satlin A, van der Geyten S, et al. on behalf of the EPAD project partners. European prevention of Alzheimer's dementia (EPAD) project: An international platform to deliver proof of concept studies for secondary prevention of dementia. J Prev Alzheimers Dis 2014;1:221 [Abstract OC7].

[55] Vradenburg G, Rubinstein E, Lovestone S, Egan C, Duggan C, Lappin DR, et al. The global Alzheimer platform: a building block for accelerating clinical R&D in AD. J Prev Alzheimers Dis 2014; 1:221 [Abstract P2–15].

Clinical Trials

NCT00231946, ALADDIN study–phase 2I: antigonadotropin-leuprolide in Alzheimer's disease drug investigation (VP-AD-301). For status see: https://clinicaltrials.gov/ct2/show/NCT00231946. Accessed January 19, 2015.

NCT00432081, Effect of indomethacin on the progression of Alzheimer's disease. For status, see: https://clinicaltrials.gov/ct2/show/NCT00432081. Accessed January 19, 2015.

NCT00762411, Effects of LY450139, on the progression of Alzheimer's disease as compared with placebo (IDENTITY-2). For status, see: https://clinicaltrials.gov/ct2/show/NCT00762411. Accessed January 19, 2015.

NCT01760005, Dominantly Inherited Alzheimer Network trial: an opportunity to prevent dementia. a study of potential disease-modifying treatments in individuals at

risk for or with a type of early onset Alzheimer's disease caused by a genetic mutation (DIAN-TU). For status see: https://www.clinicaltrials.gov/ct2/show/NCT01760005. Accessed January 19, 2015.

NCT01767311, A study to evaluate safety, tolerability, and efficacy of BAN2401 in subjects with early Alzheimer's disease. For status see: https://clinicaltrials.gov/ct2/show/NCT01767311. Accessed December 16, 2015.

NCT01900665, Progress of mild Alzheimer's disease in participants on solanezumab versus placebo (EXPEDITION 3). For status see: https://clinicaltrials.gov/ct2/show/NCT01900665. Accessed January 19, 2015.

NCT02008357, Clinical trial of solanezumab for older individuals who may be at risk for memory loss (A4). For status see: https://clinicaltrials.gov/ct2/show/NCT02008357. Accessed January 19, 2015.