Research Paper

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

Peter Lockwood,[1,4] Wayne Ewy,[1] David Hermann,[2] and Nick Holford[3]

Objective. Clinical trial simulation (CTS) was used to select a robust design to test the hypothesis that a new treatment was effective for Alzheimer's disease (AD). Typically, a parallel group, placebo controlled, 12-week trial in 200–400 AD patients would be used to establish drug effect relative to placebo (i.e., Ho: Drug Effect = 0). We evaluated if a crossover design would allow smaller and shorter duration trials.

Materials and Methods. A family of plausible drug and disease models describing the time course of the AD assessment scale (ADAS-Cog) was developed based on Phase I data and literature reports of other treatments for AD. The models included pharmacokinetic, pharmacodynamic, disease progression, and placebo components. Eight alternative trial designs were explored via simulation. One hundred replicates of each combination of drug and disease model and trial design were simulated. A 'positive trial' reflecting drug activity was declared considering both a dose trend test ($p < 0.05$) and pair-wise comparisons to placebo ($p < 0.025$).

Results. A 4×4 Latin Square design was predicted to have at least 80% power to detect activity across a range of drug and disease models. The trial design was subsequently implemented and the trial was completed. Based on the results of the actual trial, a conclusive decision about further development was taken. The crossover design provided enhanced power over a parallel group design due to the lower residual variability.

Conclusion. CTS aided the decision to use a more efficient proof of concept trial design, leading to savings of up to US\$4M in direct costs and a firm decision 8–12 months earlier than a 12-week parallel group trial.

KEY WORDS: Alzheimer's disease; clinical trial simulation; pharmacokinetic pharmacodynamic model; trial design.

INTRODUCTION

Clinical trial simulation (CTS) may be used to assess how different design and drug factors may affect trial performance. These factors may be controllable trial design properties, e.g., the doses studied, the duration of treatment, or uncontrollable drug characteristics, such as its pharmacokinetic or pharmacodynamic models and parameters. 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. Provided there are specific decision rules for determining that a particular trial was positive, or for judging an estimate to be sufficiently accurate, CTS can provide a rational basis for

making decisions about the trial design and quantitating how effectively the design can answer the study objectives (1,2). 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 (3).

A CTS study is presented which guided the design of a proof of concept (POC) trial for an M1 muscarinic agonist (CI-1017) being developed for the symptomatic treatment of Alzheimer's disease (AD) (4). M1 receptors are abundant in the hippocampus and cortex playing an important role in learning and memory and are an attractive target for cognitive enhancement in AD (5,6). Direct acting muscarinic agonists may also be effective in modulating tissue levels of the amyloid β peptide, which in its aggregated form is highly toxic to neurons and is a major constituent of senile plaques characteristic of the neuropathological diagnosis of AD (7).

Because the mechanism of action of CI-1017 was untested clinically, the principle objective of the clinical study was to ascertain whether CI-1017 improved cognitive

[1] Pfizer Global Research and Development, Ann Arbor, Michigan, USA.

[2] deCODE genetics, Brighton, Michigan, USA.

[3] Department of Pharmacology and Clinical Pharmacology, School of Medicine, University of Auckland, Auckland, New Zealand.

[4] To whom correspondence should be addressed. (e-mail: peter.lockwood@pfizer.com)

performance at least as fast and as well as tacrine—a commercially available product for this indication. This would be considered proof-of-concept (POC). At this stage of development, while it was not essential to obtain accurate estimates of the magnitude of drug effect, it was necessary to obtain information that would enable an early decision about whether to continue investing resources in the development of the compound. This distinction in key objectives and awareness that the POC study design was for internal decision making, prompted considerations of study designs that otherwise might not be considered suitable were it a pivotal registration study.

Preclinical pharmacological studies in rats and mice which were deficient in performing spatial memory tasks (i.e., the ability to find a hidden platform in a water maze (8)) as a consequence of genetic modification or lesioning of the nucleus basalis, responded to treatment with CI-1017 but demonstrated a U-shaped dose-response (DR) relationship. As doses were increased, the latency time to find a hidden platform decreased, but at higher doses the latency time increased. A similar outcome was observed in another study that investigated reversal of scopolamine-induced impairment of vigilance in rhesus monkeys, as measured by a continuous performance task. This U-shaped phenomenon has also been described in a clinical setting by Soncrantt *et al.*, who reported that the cholinergic agonist arecoline (an alkaloid of the betel nut with agonist activity at M1 and M2 receptors) improved the ability to remember verbally presented words, in a small group of subjects with probable or possible Alzheimer's disease (9). When the results were averaged across patients a U-shaped DR curve was observed. Based on these preclinical and clinical data, it was considered plausible that CI-1017 might also have a U-shaped DR profile and therefore a secondary trial objective was to characterize the dose-response.

Typically, effectiveness trials in AD are based on a parallel group design with two to four treatment groups plus a placebo group powered to detect a three-point improvement in the Alzheimer's Disease Assessment Scale-Cognitive subscale (ADAS-Cog) score after a minimum of 12 weeks of treatment (10–12). The ADAS-Cog is an objective test that evaluates memory, attention, reasoning, language, orientation and praxis (maximum score, 70); a decrease in score indicates improvement (13). Assuming a three-point treatment effect size from placebo with a standard deviation of 5.7, a parallel group design requires approximately 80 subjects per dose group to have 90% power (based on a two sided test and 5% significance level). It has also been consistently demonstrated that cholinesterase inhibitors may take up to 12 weeks to fully reflect the response to a given (14,15) dose and therefore studies designed to demonstrate the effectiveness of a pharmacological agent are at least of this duration.

Non-human toxicology studies for CI-1017 permitted up to 12 weeks of treatment and a POC study of this duration was proposed for CI-1017. Human pharmacokinetic and safety data from the Phase 1 single and multiple dose studies indicated that 25 mg TID was the maximum dose with tolerable adverse events in healthy elderly patients. Other doses available for study were 2, 5, 10 and 15 mg.

A population pharmacokinetic pharmacodynamic analysis of tacrine clinical studies estimated a 3-unit improvement on the ADAS-Cog at 80 mg with the time to reach 50% of this drug effect (equilibration half-time) of about 2 weeks (14). This implied that the full drug effect (i.e., effect at pharmacodynamic equilibrium) would be apparent by 12 weeks of treatment. As it was not a requirement that the full drug effect be measured to demonstrate that the treatment was superior to placebo, designs that studied the change in ADAS-Cog after treating Alzheimer's patients for less than 12 weeks were considered. Subsequently, a clinical trial simulation study was undertaken to compare how well different designs of approximately equal cost, could meet the POC objectives for a range of drug candidates that differed in their pharmacodynamic properties.

Simulation Study Objectives

The primary objectives of the simulation study were to compare the power of the different study designs to detect a treatment effect of a specified size and to compare design performance to differentiate a DR pattern that was monotonic, from one that was U-shaped. A secondary objective was to evaluate the bias in the drug effect size estimate relative to that assumed at pharmcodynamic equilibrium.

As the true underlying effect for this novel treatment was not known, four theoretical DR curves were considered, each reflected by a unique dose-response model of the drug effect. The curves had either a monotonic (linear, hyperbolic, or sigmoidal) or U-shaped DR relationship and an effect size of 3-unit at the best dose in the range up to 25 mg. Alternatives were also considered for the time to reach the pharmacodynamic equilibrium state, being either "fast" (50% of the full drug effect by 3 days) or "slow" (similar to tacrine and with 50% of the full drug effect reached at 2 weeks). Based on the pharmacokinetic convention that five half-lives approximate a steady state, these drugs would reach the pharmacodynamic equilibrium state at about 2 and 10 weeks, respectively. A target sample size was set at 60 based on a preliminary power analysis for a Latin Square design and budget considerations.

MATERIALS AND METHODS

The Simulation Model

Population Pharmacodynamic Model

A population pharmacodynamic model, relating plasma concentrations to ADAS-Cog scores was used to simulate realistic patient responses (14,16). Drug effect was assumed to be symptomatic rather than disease modifying (17). The model and its parameters were based on 5,263 ADAS-Cog measurements obtained from 909 patients recruited in French and American tacrine studies (14). The ADAS-Cog score at any time during the study (i.e., $S(t)$), was a function of a linear combination of sub-models (Eq. (1)) that included a baseline value (S_0), a linear time course for untreated disease progression ($\alpha \cdot t$), a placebo effect ($PD(C_{ep})$) and drug effect ($PD(C_{eA})$). The residual error (ε) was assumed to be normally distributed with variance σ^2.

$$S(t) = S_0 + \alpha \cdot t + PD(C_{eP}) + PD(C_{eA}) + \varepsilon \qquad (1)$$

The delay in reaching a pharmacodynamic equilibrium state was characterized using the effect compartment model which assumes the drug effect is dependent on the effect site concentrations which differ from systemic concentrations prior to the attainment of pharmacokinetic equilibrium (18). Accumulation of drug at the effect site was dependent on the plasma concentration that was predicted by the CI-1017 pharmacokinetic model, and the effect compartment equilibration half-time ($t_{1/2eq}$).

Placebo Response Model (PD(C_{ep}))

Patients entering an Alzheimer's trial may demonstrate a placebo response. Therefore, a placebo component was included in the simulation model. Based on the placebo model described by Holford and Peace (16), a placebo response ($ADAS\text{-}Cog_p(t)$) was assumed, that would develop over the early days of treatment and subsequently fade to zero after about 6 weeks. Mathematically, the time course of placebo response can be described by Eq. (2) with parameters

Keq_p, the rate constant defining the onset rate of the placebo effect (days^{-1}); Kel_p, the rate constant defining the offset rate of the placebo effect (days^{-1}), and β_p, a scaling parameter defining the size of the placebo effect. The time in days after the initiation of treatment is represented by the independent variable *t*.

$$ADAS - Cog_p(t) = \frac{\beta_p \cdot Keq_p}{\left(Keq_p - Kel_p\right)} \cdot \left(e^{-Kel_p \cdot t} - e^{-Keq_p \cdot t}\right)$$

(2)

The onset half-life of the placebo effect is related to Keq_p by Eq. (3).

$$t_{1/2eqm} = \frac{\ln(2)}{Keq_p}$$

(3)

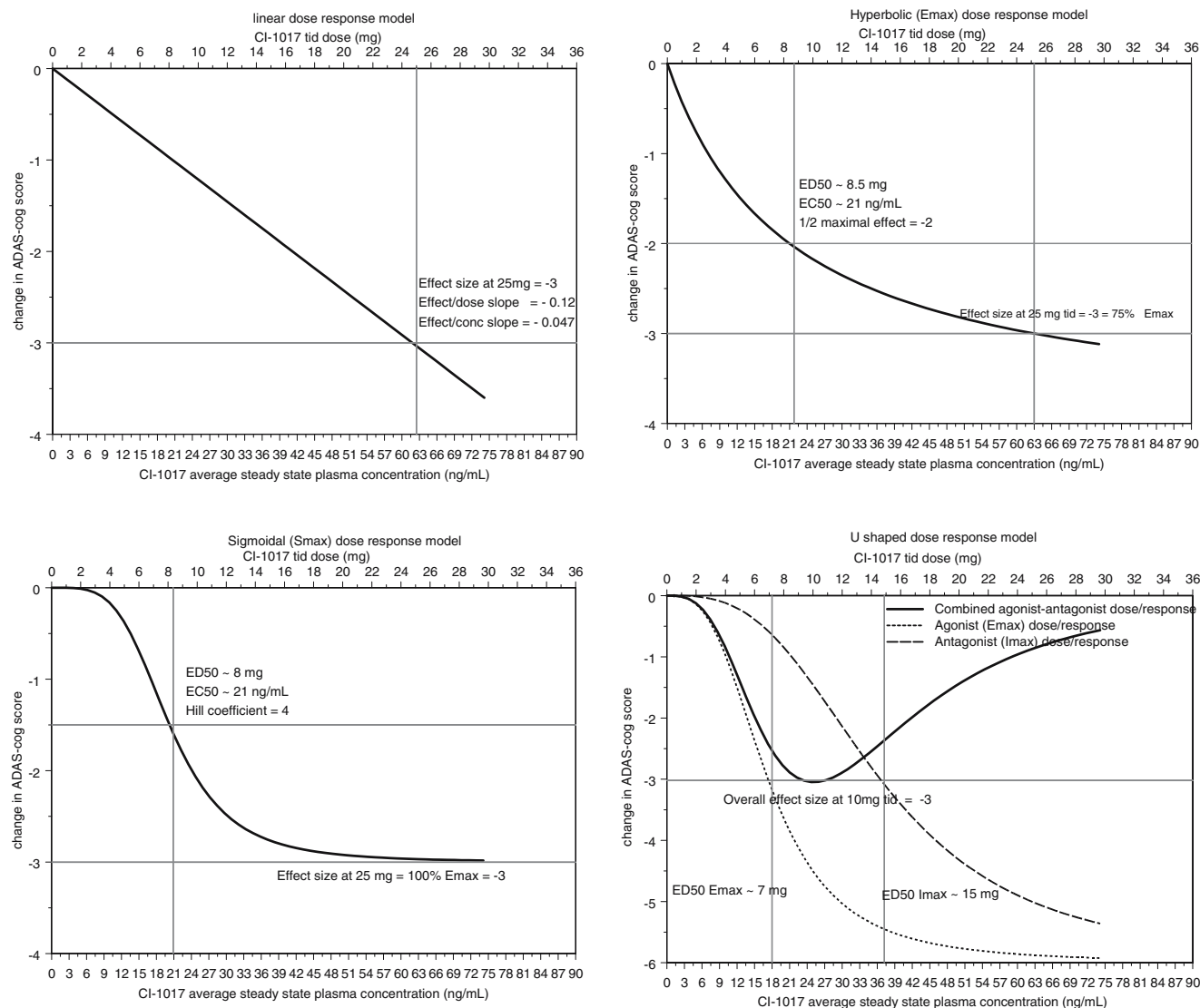The offset half-life of the placebo effect is related to Kel_p in a similar fashion.

Fig. 1. Drug effect models considered in simulations study. Parameters characterizing the model are displayed in the individual panels.

Table I. Parametric Form of Models Describing Active or Inactive Drug Effect. Model Parameters are Described in Table II

Model Type	Description	Parametric Form
Inactive	No effect	$PD_{inactive}(C_{eA}) = 0 \cdot C_{eA}$
Active	Linear	$PD_{Emax}(C_{eA}) = \beta_A \cdot C_{eA}$
	E_{max} (hyperbolic)	$PD_{Emax}(C_{eA}) = \dfrac{E_{max} \cdot C_{eA}}{EC_{eA50} + C_{eA}}$
	S_{max} (sigmoidal)	$PD_{Smax}(C_{eA}) = \dfrac{E_{max} \cdot C_{eA}^n}{EC_{eA50}^n + C_{eA}^n}$
	U-shaped [A, B]	$PD_{Ushape}(C_{eA}) = E_{max\,U} \cdot \left(\dfrac{C_{eAU}^n}{EC_{eAU_{50}}^{\,n} + C_{eAU}^n} - \dfrac{IC_{eaU}^n}{IC_{eaU_{50}}^{\,n} + IC_{eaU}^n} \right)$

[A] The U-shaped dose-response pattern was modeled as the difference of two E_{max} functions with the same E_{max} and Hill coefficient (agonist–antagonist model).
[B] C_{eA}, CI-1017 drug concentration at effect site for monotonic concentration response relationship.
C_{eaU}, CI-1017 drug concentration at effect site mediating pure agonist effect.
IC_{eaU}, CI-1017 drug concentration at effect site mediating antagonist (inhibitory) effect.

Drug Response Model (PD(C_{eA}))

Five different drug response models were used. One model characterized the drug as inactive (no effect) while the other four described a linear, hyperbolic, sigmoidal or U-shape. Mean parameter values for the active concentration response relationships were calculated based on lowering the mean ADAS-Cog score by three points at the best dose in the testable range. This was at 25 mg for monotonic patterns and 10 mg for the U-shape pattern. Figures and parametric forms of the drug effect models are displayed in Fig. 1 and Table I, respectively. The parameters characterizing drug potency for the linear and non-linear models (i.e., slope or ED 50 [the dose at which 50% of the maximum drug effect was achieved] were converted to an average steady state concentration in a 65-year old non-smoking population.

To reflect between-patient random variation, the baseline ADAS-Cog score (S_0), rate of score change (α), onset half-life of placebo effect ($t_{1/2eqp}$), offset half-life of placebo effect ($t_{1/2elp}$), placebo effect magnitude (β_p), and active drug

Table II. Parameter Values Characterizing Disease Progression, Placebo Effect, and Drug Effect Components used in Simulation Model

Parameter	Description	Units	Population Estimate	Percentage CV
S_0	Baseline ADAS-Cog score	ADAS-Cog units	30	30
α	Rate of change of ADAS-Cog score, due to disease progression	U/day	0.0164	30
$t_{1/2eqp}$	Placebo effect onset half-life mediator	Days	6	30
$t_{1/2elp}$	Placebo effect offset half-life	Days	7	30
β_p	Magnitude of placebo effect	ADAS-Cog	−3	30
$t_{1/2eqa}$	Equilibration half-life (slow)	Days	16	30
	Equilibration half-life (fast)		2.8	30
β_A	Potency of active drug (linear model)	ADAS-Cog*ml/ng	−0.047	30
E_{max}	Maximum drug effect (hyperbolic model)	ADAS-Cog units	−4	NA
	Maximum drug effect, (sigmoidal model)	ADAS-Cog units	−3	NA
	Maximum drug effect for U-shaped model	ADAS-Cog units	−6	NA
EC_{eA50}	effect compartment concentration producing 50% of E_{max}	ng/ml	21	30
EC_{eAU50}	Effect compartment concentration producing 50% of Emax for agonist component of U-shape model	ng/ml	18	30
IC_{eAnU50}	Effect compartment concentration producing 50% inhibition of Emax for inhibitory component (antagonist) of U shape model	ng/ml	38	30
N	Hill coefficient (sigmoidal model)	Unitless	4	30
	Hill coefficient (U shape model)	Unitless	3	30
σ	Residual error standard deviation	ADAS-Cog units	4	NA

NA = Not assigned.

Table III. CI-1017 Population Pharmacokinetic Parameter Estimates

Parameter	Units	Value	Population Predicted Variance	FBOV
Clearance	l/h	94.5	32	11
Central volume	l	172	45	21
Inter compartmental clearance	l/h	31.8	75	43
Peripheral volume	l	222	71	41
Absorption rate constant	h^{-1}	4.81	103	ND
Lag time	h	0.322	27	ND

*FBOV = Fraction of population variance attributable to between occasion variability.

equilibration half-life ($t_{1/2eqa}$), were simulated from univariate independent log-normal distributions. The coefficient of variation (CV) for these distributions is shown as percentage CV in Table II.

Disease progression, placebo response, and drug effect parameter values are displayed in Table II. Population variability parameters were arbitrarily set to 30% based on discussions with consultants and values reported by Holford and Peace (14). The drug effect equilibration half-time and placebo response reported were associated with the different study populations and the parameter estimates selected for these components in this simulation study reflected an American population (14).

Pharmacokinetic Model

Single doses ranging from 0.25 to 150 mg and multiple doses of 2, 10, 25, and 50 mg q6h were rapidly absorbed. The time of maximal plasma concentrations (t_{max}) occurred approximately 1 h after oral administration. With multiple-dose administration of CI-1017, maximum plasma drug concentrations (C_{max}) and total exposures (AUC) increased in greater proportion relative to the increase in dose over a dose range of 2 to 50 mg. Multiple doses of 2, 10, and 25 mg administered q6h were generally well-tolerated.

Population pharmacokinetic parameters and estimates of their variability were derived from fitting the Phase 1 data using NONMEM and are displayed in Table III (19). Plasma

concentrations of CI-1017 were described by a two-compartment model with first order input and a lag-time. The relative bioavailability of each dose was estimated in comparison to the nominal standard dose of 25 mg. Clearance (CL/F) was dependent on age (years) and smoking status (Eq. 4).

$$CL = 94.5 \cdot e^{-0.0135 \cdot (age-40)} \cdot smk \qquad (4)$$

If the subject was a smoker, the variable smk was equal to 1.5, otherwise 1. The between-subject random effects for some pharmacokinetic parameters (clearance, central volume, inter-compartmental clearance, and peripheral volume) were correlated. The correlation was incorporated into the simulation model by using a multivariate normal distribution from the estimated variance–covariance matrix obtained from NONMEM. The estimated between occasion variability (BOV) in pharmacokinetics was also incorporated into the simulation model (20). At each 7-day occasion, random normal distributions with mean 0 and variance σ_{BOV}^2 were used to generate new subject specific random effects values for BOV, which were then added to the subject's pharmacokinetic parameters.

Drop Out Model

A 1% weekly dropout rate was assumed based on an informal review of past experience in these types of trials. This was implemented in the simulation model as a survival model where the expected percentage of patients remaining in the trial at time = t days, $S_v(t)$, is described by Eq. (5).

$$S_v(t) = e^{-0.00145 \cdot t} \qquad (5)$$

Trial Designs

Eight trial designs were evaluated, including Latin Square, incomplete block, and parallel group, as well as two composite designs that included both crossover and parallel group arms. Each design had approximately equal sample size (60 total), as a surrogate for trial cost. No patient received non-placebo treatment for more than 12 weeks—the total trial length was 12 weeks for all designs except for number 8, which was 16 weeks. The key characteristics of these designs are shown in Table IV. Table V shows the

Table IV. Trial Designs Evaluated in Simulation Study

Design Number	Design Description	Number of Sequences	Subjects per Sequence	Number of Treatments Periods	Period Length (weeks)	Measurements per Period
1	6 × 6 Latin Square	6	10	6	2	1
2	6 × 3 Incomplete block	6	10	3	4	2
3	Parallel group	6	10	1	12	6
4	6 × 4 Incomplete block	6	10	4	3	1
5	6 × 3 Incomplete block with two parallel groups			Seq 1–6: 3	Seq 1–6: 4	2
		8	8	Seq 7–8: 1	Seq 7–8:12	6
6	4 × 4 Latin Square	4	15	4	3	1
7	4 × 4 Latin Square with two parallel groups			Seq 1–4: 4	3	1
		6	10	Seq 5–6: 1	12	6
8	4 × 4 Latin Square	4	15	4	4	2

Table V. Dosing Sequences in Design Number 1

Design Number 1: 6 × 6 Latin Square; TID Dose (mg)					
Sequence number 1	P	25	2	15	5	10
Sequence number 2	2	P	5	25	10	15
Sequence number 3	5	2	10	P	15	25
Sequence number 4	10	5	15	2	25	P
Sequence number 5	15	10	25	5	P	2
Sequence number 6	25	15	P	10	2	5

dosing sequences for the six groups in design number 1, a six period, six-sequence group Latin Square that is balanced for pair-wise treatment sequence—a William's design (21). Design number 6–8 were based on a similarly defined 4 × 4 Latin Square, using 2, 10, and 25 mg doses and placebo. The lower doses (2, 10 mg) were selected to approximate the ED_{10} and ED_{50} while the 25 mg dose was selected to achieve a 3-unit change in the ADAS-Cog for the assumed population average monotonic dose-response relationships. For the U-shaped dose-response relationship, 10 mg defined the population average peak effect, while the bordering doses of 2 and 25 mg doses mediated a small effect (less than 25% of E_{max} for the agonist–antagonist model).

Data Evaluation Methods

An analysis of variance (AOV) model appropriate to each design was used to analyze the simulated data. For each trial objective, a decision rule was defined to translate a trial's analysis results into an outcome from which the percentage correct was calculated (over the 100 replicate trials). The analysis methods employed for each of the trial objectives are detailed more explicitly as follows.

Primary Objective A: Does the drug work?

The AOV was used to test the null hypothesis of no drug effect on ADAS-Cog. Rejection of the null hypothesis constituted a "positive study" finding, which would be judged as "correct" for all data models except the no effect model, for which "not positive" was correct. The specific AOV model depended upon the design, but the decision rule for declaring a "positive study" was based on a 2 degrees of freedom (linear and quadratic) dose trend test at the two-tailed 5% significance level together with at least one dose declared statistically better than placebo ($p < 0.025$, one tail).

Primary Objective B: Is the shape monotonic or U-shaped?

After exploring various parametric and non-parametric approaches, a simple but robust criterion was adopted for deciding whether a trial's data supported a monotonic or U-shaped pattern (within the tested dose range). A two-stage approach was used, initiated with a test for activity that was identical to the procedure used in primary objective A, except for using a higher type 1 error setting (0.1 one tail rather than 0.025 one tail). As this was not a pivotal registration study and was to be used for internal decision-making, relaxing of the critical value for rejecting the null hypothesis and increasing the Type I error to 0.1 for the shape classification algorithm was considered justified. For a non-positive trial (by this relaxed standard), the response

Table VI. Estimated Power for Detecting Activity (%), Averaged over the Four Dose-response Patterns

Design Number	1	2	3	4	5	6	7	8
Fast	95	64	41	56	64	86	80	96
Slow	48	40	30	31	41	63	54	81

pattern was classified as 'flat' (i.e., essentially as 'no information regarding shape'); otherwise an inference was made between monotonic and U-shaped based on the pattern of the estimated dose group means from the AOV. For the four-dose group designs, monotonicity was declared if the highest dose group had the best mean outcome; otherwise a U-Shaped pattern was declared. For the six-group designs, monotonicity was declared if either of the two highest dose groups had the best mean.

Secondary Objective: What is the Accuracy of the Pharmacodynamic Equilibrium Effect Estimate?

An effect size estimate within one point (33%) of the true steady state effect size was considered "correct" for the purposes of the simulation study. The simulated estimate at the 25 mg dose was used for the monotonic DR patterns (regardless of the outcome for objective number 2), or the 10 mg dose when the true pattern was U-shaped.

Simulation Methods

One hundred trial replications were simulated using Pharsight Trial Simulator (22) for the different study designs listed in Table IV, for each of the dose-response patterns and for both the *slow* and *fast* drug types. The residual error, the effect size at the best dose, and the number of subjects were held constant, while the period length, number of doses, and number of measurements per period were varied. The data from each trial were analyzed to obtain and score the conclusions for each trial objective. The percentages of correct trials were tabulated for review, leading to the design recommendations described below.

The Clinical Study

Based on the results of the simulations, the recommended design was carried out in an Alzheimer's population. Both the 2- and 4-week measurements within each dose were

Table VII. Percent of 100 Trials (power) that Detected a Drug Effect for Slow Acting Drug for Design Number 6, 7 and 8

Design Number Dose-response shape	8	7	6
Linear	84	41	51
E_{max}	88	58	67
S_{max}	96	75	85
U-shape	57	40	49
Average	81	54	63

Design number 6: 4 × 4 Latin Square, 3 weeks per treatment.
Design number 7: 4 × 4 Latin Square with two parallel groups.
Design number 8, 4 × 4 Latin Square, 4 weeks per treatment.

Table VIII. Percent of 100 Trials (power) that Correctly Identified Dose-response Shape for Slow Acting Drug for Design Number 6, 7 and 8

Design Number Dose-response shape	8	7	6
Linear	96	69	72
E_{max}	84	62	74
S_{max}	96	83	89
U-shape	45	34	39
Average	80	62	69

used in the analysis of the ADAS-Cog subscale. The changes from baseline in total score on the ADAS-Cog were analyzed using a mixed-model analysis of variance that was similar to the method used in the simulation. The model contained fixed effect terms for treatment sequence, period, carryover effect, baseline value, and dose. Random effects were included to model the correlation between measurements within a patient, and to model the correlation between the 2-and 4-week measurements within each dose. A two-degree of freedom linear-quadratic trend test was performed to investigate dose-response (5%, two-tail). The study was to be considered positive if both the trend test (5%, two-tail) and the improvement of ADAS-Cog relative to placebo, at one or more active doses (2.5%, one-tail) were significant.

RESULTS

Simulation Study Results

For detecting drug activity, all designs correctly declared the no effect drug candidate as "not positive" with an error rate of about 2.5% (observed rates based on 100 simulated trials ranging from 1% to 4%). For each design, the percentage of positive trials averaged over the four dose-response patterns is displayed in Table VI, for the slow and fast acting drug types.

For each drug type, the best designs were 1, 6, 7, and 8, with design number 8 (4 × 4 Latin Square, 4-week treatment periods) being the best for both drug types. The parallel group design (number 3) was the poorest performer for both drug types. Only design 8 reached the 80% power level on average over the four dose-response patterns with the sample sizes used. Across designs, the estimated power was always higher for the fast acting drug. Therefore in order to simplify

Table IX. Percent of 100 Trials (Power) that Estimated the Effect Size for the Slow Acting Drug at 25 mg (monotonic) or 10 mg (U-shape) to be within +/− 33% of the True Effect

Design Number Dose-response shape	8	7	6
Linear	42	41	42
E_{max}	47	51	55
S_{max}	75	68	65
U-shape	14	28	34
Average	45	47	49

Table X. Analysis of ADAS Cognitive Results (Intent-to-Treat) from a 16-week, Randomized, Double-blind, Placebo-Controlled, Crossover, Multi-Center Study of CI-1017 in Patients with Mild to Moderate Alzheimer's Disease

TID Dose (mg)	Difference from Placebo (95% CI)
2	−0.46 (−1.14–0.23)
10	−0.06 (−0.75–0.62)
25	−0.68 (−1.37–0.01)

the description of further results the design performance for the slow acting drug only is reported, as this will characterize trial performance for the least optimistic scenario. Table VII partitions the average power to detect a drug effect into the results for each of the active drug profiles. Table VIII displays the estimated power to correctly identify the dose-response shape. For all shapes, design 8 performed as well or better than designs 6 and 7.

Table IX displays the power to correctly estimate the true effect size (to within +/− 33% of the true effect). As expected the overall level of performance is low. Averaged over the four DR models, designs 6, 7 and 8 are similar. For all the objectives, the power of any design was always lower for the U-shaped drug type compared to the drug type with monotonic dose-response characteristics.

Based on these results, design 8 was selected for the actual clinical trial. Seventy patients were randomized to one of four assigned treatment sequences. Each treatment sequence consisted of three dose levels of CI-1017 (2, 10, and 25 mg) taken three times a day (TID) and matching placebo with each dose level administered sequentially. The sequence of administration was randomly assigned. All patients were intended to receive 4-week treatment with each of the dose levels of CI-1017 and placebo.

Clinical Study Results

The results of the clinical study are displayed in Table X. The study was not considered positive since the linear-quadratic trend test of ADAS-Cog scale was not significant and none of the three doses of CI-1017 were significantly better than placebo. The root mean square error (RMSE) derived from the mixed-model analysis of variance was 2.47 ADAS-Cog units.

DISCUSSION

It is often thought that the most convincing studies that demonstrate the benefit of therapy are parallel group designs with long treatment phases, in part because the drug effect at the time an equilibrium state has been reached is considered to be the critical endpoint. Crossover designs on the other hand are considered to harbor methodological problems and confound estimates of the treatment effect under various conditions, particularly when carry-over is present (3). This paper describes a clinical trial simulation study that provided a quantitative comparison of the potential performance of different study designs in Alzheimer's disease for fast and slow acting drug types and ultimately supported the implementation of a cross-over design for a POC study. While this example is

limited to a POC trial in Alzheimer's disease, the concepts discussed are not limited to that therapeutic area and would be applicable to many Phase 2 and 3 clinical trial settings.

The simulation model was characterized by parametric descriptions of the drug pharmacokinetics and pharmacodynamics, the natural progression of the disease, and the placebo effect. Estimates of the residual error and within subject variability for the different parameters were also included in the model. The simulated data were analyzed using a polynomial-based approach for dose-trend testing. This analysis method satisfied a basic requirement that the test be appropriate in the presence of non-monotonicity of response with dose, a requirement that excluded any trend test for which the validity depends upon monotonicity. Pragmatic considerations and time constraints were also factors in the choice of the analysis method for the simulations. It was recognized that the analysis method could have some effect on the relative estimates of design performance, however it was decided not compare the performance of different dose trend tests. This would have added another dimension to what was already a complex project, potentially compromising the timely delivery of the simulation results to the development team. Holford reports between-subject variability of 208% for disease progression and approximately 100% for placebo and drug potency parameters (14) and did not estimate random effects for any other parameters. In this simulation study, parameters accounting for the between-subject variability were assigned to all fixed effect parameters with the exception of E_{max} and this partitioning was considered to result in an overall between-subject variability that approximated that reported by Holford. However, as the primary consideration was to understand the performance of a trial design using this set of assumptions, the impact of the size of the random effects was not explored. The residual variability matched that described in the literature.

Because the model and model parameters were based on data analyzed almost 15 years ago (14,16), the possibility of a change in the time course of the disease and placebo response between former and latter patient populations merits comment. In a study of 331 patients receiving a placebo treatment reported by Feldman et al. in 2005 (23), the mean 12-month decline in ADAS-Cog was 5.6 U/year (95% CI: 4.8–6.4). Stern et al. (1994), reported a decline of 4.9 points that falls within this confidence interval (24), while a number of placebo-controlled studies (25–29) have shown that the annual decline in cognition is 5–8 points with a 2-year decline of seven points reported by Sano et al. in 1997 (28). Nevertheless, even if disease progression rate was slower than that used in this simulation study, the impact would be minimal because the effect would be equal for all treatment groups and the duration of the study was short. With regard to the placebo response, it is difficult to make assertions as to whether this profile has changed over time because investigators do not usually differentiate the placebo response from the underlying disease progression.

As the principle goal of this Phase 2a study was to answer the key questions as to whether the drug had any effect at all given the target sample size, as quickly as possible, it was not considered necessary to measure the steady state treatment effect. Additionally, even though the

trial was considered a tool for internal decision-making and would not satisfy the regulatory requirements for a confirmatory study, it was believed that a positive outcome would provide suitable supportive data for registration. The 4×4 crossover design was demonstrated to be the best design among those considered for detecting activity. For the fast acting design, which produces a larger effect size sooner than the slower one, the power for detecting the effect was higher regardless of design. Therefore the design evaluation focused on the performance assuming administration of the drug type that was slower acting. The lower power that was consistently observed for the drug type with U shape characteristics reflects the fact that the mean effect size at any of the dose options was less than 3, even though the population model defined a three-point effect at 10 mg. Due to the variability in EC_{eAU50} and IC_{eAnU50} parameters specified in the agonist–antagonist simulation model, an individual's "best dose" would not necessarily be at any of the doses included in the simulation study. Consequently this would lead to an average effect size of less that three points at any one of the doses studied. This simulation approach reflects inter-subject variability to drug potency, which was considered a realistic pharmacological concept. Other methods were not considered pharmacologically realistic e.g., an agonist antagonist model that excluded inter-subject variability and only described residual variability.

While there were significant advantages to the implementation of the cross-over design in this instance, the design and analysis strategy (ANOVA) itself does have limitations, and these should be addressed when considering this type of study. For example, the design and analysis is unsuitable for evaluating beneficial or adverse treatment effects that are apparent only after 4 weeks of therapy because the observed effects would be assigned to the wrong treatment. Further, this design would be inefficient for evaluating effects that persist unchanged into subsequent periods; the effect estimates become increasingly biased as carry-over increases. Alternatively, a pharmacologically based model analysis would explicitly account for carryover in a mechanistic way.

Further development of CI-1017 was terminated based on the results of the clinical study. The residual error standard deviation of 2.47 estimated from the actual POC study was 60% of the value observed in parallel group tacrine trials and 60% of the value used in the trial simulations (four ADAS-Cog units). This implied that the study was better powered to detect a signal than the power associated with a similar sized parallel group trial and the outcome appeared consistent with results for another drug in this class. Veroff et al., reported a mean treatment difference of −1.26 ($p = 0.22$), based on an intent-to-treat (ITT) analysis in patients administered 75 mg TID of xanomoline (an M1/M4 selective muscarinic agonist), for 24 weeks. Visual inspection of a figure depicting the change in mean ADAS-Cog score at 4, 8, 12 and 24 weeks that was based on the ITT population, suggests no difference between the placebo and drug treated groups at 4 weeks and that the maximal drug effect is exerted between 8 and 12 weeks (30).

The implementation of the crossover design resulted in substantial time and cost savings relative to a traditional parallel group study. The minimum direct costs associated with a conventional 12 week, 300 patient Alzheimer's POC

study in 25–30 centers was estimated to be about US$4M (2001) and to extend over 24-month duration, from the time of enrollment of the first subject until the reporting of the results. In this example, because fewer patients and recruitment centers were required, the study was executed for one quarter of the cost (approximately US$1M) and the time from enrollment of the first patient to reporting of results was reduced to 7 months. These time-savings enabled resources to be reallocated to other development projects and the additional value of this "indirect saving" is considerable and should not be overlooked. The approach of initially determining whether the drug has any activity over a short time period, in place of estimating the drug effect after long-term treatment enables an earlier termination of a development program if a drug effect is not detected. Such an approach is consistent with the theme of developing new tools and practices to improve the critical path or product development process as described in the FDA's recent document that addresses how to get fundamentally better answers about the safety and effectiveness of new medical products (31).

The use of placebo controls in Alzheimer's disease has been the center of debate and is considered an unethical practice by some, especially if the treatment is long term. Use of the crossover design as described herein, may offer a solution to this dilemma in POC studies of acute symptomatic treatments because patients are assigned a placebo treatment for much shorter periods (32,33). Another benefit of the shorter placebo period might be reduced drop out due to lack of effectiveness and enhanced recruitment as all patients will receive study drug during the trial.

This simulation study reflects a changing attitude in drug development. Firstly, the primary objective was to obtain sufficient evidence to determine if further development of the compound was justifiable by focusing on whether any activity could be detected after 4 weeks, as opposed to estimates of the drug effect after 12 weeks. Estimating the drug effect following 12 weeks of treatment was considered the goal of a of Phase IIB dose-response study. Secondly, a semi-mechanistic PKPD approach was used for simulating realistic patient level data as a function of drug exposure and time, and this enabled a quantitative comparison of the performance of different study designs under the different scenarios. This quantitative comparison eventually convinced skeptics to move forward with an atypical design. Because the study provides a "yes or no answer" to the fundamental question "does the drug work," it can be described as a confirming type study. Sheiner has previously discussed the distinction between learning and confirming studies (34). Had the study yielded a positive outcome, analysis of the data in a stand alone mode or coupled with prior information, may have then provided the answers to more quantitative questions such as "what is the expected steady state effect size at each dose"—a "learning" type study.

In summary, these clinical trial simulation results helped the development team better understand and compare the operating characteristics of eight plausible POC trial designs. In the end the chosen design proved to be more efficient than a traditional clinical development approach leading to considerable savings in time to decision and trial costs. The results of this CTS support the continued application of modeling and simulation to aid the design of clinical studies.

REFERENCES

1. P. L. Bonate. Clinical trial simulation in drug development. *Pharm. Res.* **17**:3252–3256 (2000).
2. N. H. G. Holford, H. C. Kimko, J. P. R. Monteleone, and C. C. Peck. Simulation of clinical trials. *Ann. Rev. Pharmacol. Toxicol.* **40**:209–234 (2000).
3. M. E. Putt and B. Ravina. Randomized, placebo controlled, parallel group *versus* crossover study designs for the study of dementia in Parkinson's disease. *Control. Clin. Trials* **23**:111–126 (2002).
4. G. McKhann, D. Drachman, M. Folstein, R. Katzman, D. Price, and E. M. Stadlam *et al.*, Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA work group under the auspices of department of health and human services task force of Alzheimer's disease. *Neurology* **34**:939–944 (1984).
5. A. I. Levey. Muscarinic acetylcholine receptor expression in memory circuits: implication for treatment of Alzheimer disease. *Proc. Natl. Acad. Sci. USA.* **93**:13541–13546 (1996).
6. M. R. Emmerling, R. D. Schwarz, K. Spiegel, and M. J. Callahan. New perspectives on developing muscarinic agoists for Alzheimer's disease. *Alzheimers Dis.* **2**:187–194 (1997).
7. D. J. Selkoe. Alzheimer's disease: a central role for amyloid. *J. Neuropathol. Exp. Neurol.* **43**:438–447 (1994).
8. R. G. M. Morris. Development of a water maze procedure for studying spatial learning in the rat. *J. Neurosci. Methods* **11**:47–60 (1984).
9. T. T. Soncrant, K. C. Raffaele, S. Asthana, A. Berardi, P. P. Morris, and J. V. Haxby. Memory improvement without toxicity during chronic low dose intravenous arecoline in Alzheimer's disease. *Psychopharmacology* **112**:421–427 (1993).
10. N. Qizibash, A. Whitehead, J. Higgans, G. Wilcock, L. Schneider, and M. Farlow. Cholinesterase inhibition for Alzheimer disease. A meta-analysis of the tacrine trials. *JAMA* **280**:1777–1782 (1998).
11. A. Whitehead, C. Perdomo, R. D. Pratt, J. Birks, G. K. Wilcock, and J. G. Evans. Donepezil for the symptomatic treatment of patients with mild to moderate Alzheimer's disease: a meta-analysis of individual patient data from randomized controlled trials. *Int. J. Geriatr. Psychiatry* **19**:624–633 (2004).
12. N. Trinh, J. Hoblyn, S. Mohanty, and K. Yaffe. Efficacy of cholinesterase inhibitors in the treatment of neuropsychiatric symptoms and functional impairment in Alzheimer disease. A meta-analysis. *JAMA* **289**:210–216 (2003).
13. W. G. Rosen, R. C. Mohs, and K. Davis. A new rating scale for Alzheimer's disease. *Am. J. Psychiatry* **141**:1356–1364 (1984).
14. N. H. G. Holford and K. E. Peace. Results and validation of a population pharmacodynamic model for cognitive effects in Alzheimers patients treated with tacrine. *Proc. Natl. Acad. Sci.* **89**:11471–11475 (1992).
15. A. Burns, M. Rossor, J. Hecker, S. Gauthier, H. Petit, H. J. Moller, S. L. Rogers, and L. T. Freidhoff and the International Donepezil Study Group .The effects of Donepezil in Alzheimer's disease—results from a multinational trial. *Dement. Geriatr. Cogn. Disord.* **10**:237–244 (1999).
16. N. H. G. Holford and K. E. Peace. Methodologic aspects of a population pharmacodynamic model for cognitive effects in Alzheimers patients treated with tacrine. *Proc. Natl. Acad. Sci.* **89**:11466–11470 (1992).
17. P. L. Chan and N. H. G. Holford. Drug treatment effects on disease progress. *Ann. Rev. Pharmacol. Toxicol.* **41**:625–659 (2001).
18. N. H. G. Holford and L. B. Sheiner. Understanding the dose–effect relationship: clinical application of pharmacokinetic and pharmacodynamic models. *Clin. Pharmacokinet.* **6**:429–453 (1981).
19. L. S. Beal and L. B. Sheiner. *NONMEM Users Guides, part I–VIII*, University of California, San Francisco, 1996.
20. M. O. Karlsson and L. B. Sheiner. The importance of modeling inter-occasion variability in population pharmacokinetic analyses. *J. Pharmacokinet. Biopharm.* **21**:735–750 (1993).
21. B. Jones and M. G. Kenward. *Design and Analysis of Cross-Over Trials*, Chapman & Hall, London, 1994.
22. Pharsight Trial Simulator Version 2.1. 2000. Pharsight Corporation Mountain View California, USA.

23. H. Feldman, B. Van Baelen, S. M. Kavanagh, and E. L. Koen. Torfs, MSc§ Cognition, function, and caregiving time patterns in patients with mild-to-moderate Alzheimer disease—a 12-month analysis. *Alzheimer Dis. Assoc. Disord.* **19**:29–36 (2005).

24. R. G. Stern, R. C. Mohs, and M. Davidson. A longitudinal study of Alzheimer's disease: measurement, rate, and predictors of cognitive deterioration. *Am. J. Psychiatry* **151**: 390–396 (1994).

25. P. S. Aisen, K. L Davis, J. D. Berg, K. Schafer, K. Campbell, R. G. Thomas, M. F. Weiner, M. R. Farlow, M. Sano, M. Grundman, and L. J. Thal. A randomized controlled trial of prednisone in Alzheimer's disease. Alzheimer's disease cooperative study. *Neurology* **54**:588–593 (2000).

26. P. S. Aisen, K. A. Schafer, M. Grundman, E. Pfeiffer, M. Sano, K. L. Davis, M. R. Farlow, S. Jin, R. G. Thomas, and L. J. Thal. Effects of rofecoxib or naproxen *vs* placebo on Alzheimer disease progression: a randomized controlled trial. *JAMA* **289**:2819–2826 (2003).

27. S. A. Reines, G. A. Block, J. C. Morris, G. Liu, M. L. Nessly, C. R. Lines, B. A. Norman, and C. C. Baranak. Rofecoxib: no effect on Alzheimer's disease in a 1-year, randomized, blinded, controlled study. *Neurology* **62**:66–71 (2004).

28. M. Sano, C. Ernesto, R. G. Thomas, M. R. Klauber, K. Schafer, M. Grundman, P. Woodbury, J. Growdon, C. W. Cotman, E. Pfeiffer, L. S. Schneider, and L. J. Thal. A controlled trial of selegiline, alpha-tocopherol, or both as treatment for Alzheimer's disease. The Alzheimer's disease cooperative study. *N. Engl. J. Med.* **336**:1216–1222 (1997).

29. L. J. Thal, M. Calvani, A. Amato, and A. Carta. A 1-year controlled trial of acetyl-l-carnitine in early-onset AD. *Neurology* **55**:805–810 (2000).

30. A. E. Veroff, N. C. Bodick, W. W. Offen, J. J. Sramek, and N. R. Cutler. Efficacy of xanomeline in Alzheimer's disease: cognitive improvement measured using the computerized neuropsychological test battery (CNTB). *Alzheimer Dis. Assoc. Disord.* **12**:304–312 (1998).

31. US Department of Health and Human Services, Food and Drug Administration. Challenge and opportunity on the critical path to new medical products (2004). http://www.fda.gov/oc/initiatives/criticalpath/whitepaper.html.

32. P. J. Whithouse, R. Arizaga, H. Brodaty, and S. Gauthier. Placebo in clinical trials in Alzheimer disease: an international discussion. *Alzheimer Dis. Assoc. Disord.* **13**:121–123 (1999).

33. D. Knopman, J. Kahn, and S. Miles. Clinical research designs for emerging treatments for Alzheimer's disease. Moving beyond placebo controlled trials. *Arch. Neurol.* **55**:1425–1429 (1998).

34. L. B. Sheiner. Learning *versus* confirming in clinical drug development. *Clin. Pharmacol. Ther.* **6**:275–291 (1997).