Published in final edited form as:

Curr Alzheimer Res. 2012 December ; 9(10): 1174–1181.

Was phenserine a failure or were investigators misled by methods?

Robert E. Becker^{1,2} and Nigel H. Greig²

¹Aristea Translational Medicine Corp., Freeport, ME, 04078, USA

²Drug Design and Development Section, Laboratory of Neurosciences, Intramural Research Program, National Institute on Aging, National Institutes of Health, Baltimore MD, 21224, USA

Abstract

Over 200 Alzheimer's disease (AD) drug candidates have failed in development, and other neuropsychiatric trials have had their validity compromised. Studies suggest that methodological errors can be a source for these compromises and failures. We gained access to documentation for phenserine, an experimental AD drug that reached phase III clinical trials. The 06 Phase III trial was cited by the developers as grounds for their abandonment of the development. We compared evidence for interventions to control methodological errors and grounds for moving through phases of drug development to 40 other randomly selected AD developments we had studied. We analyzed methods and conditions of the 06 phenserine clinical trial, for biases able to account for its abandonment during development. The phenserine development failed to control error sources able to affect the outcomes. There are statistically significant relationships in the 06 clinical trial between outcomes at research sites and levels of variance, placebo group improvements and other factors. We conclude that phenserine was abandoned, at least in part, due to a clinical trial invalidated by relationships among its methods and outcomes.

Keywords

neuropsychiatric drug development; placebo group improvement; variance; validity; phenserine; scientific practices in drug development; Alzheimer's disease; clinical translation

Introduction

All scientists face the uncertainty that an outcome from an experiment may result from some influence other than the independent variable. Unlike investigators who carry out laboratory based experiments, Alzheimer's disease (AD) drug development investigators cannot readily assure that clinical trial methods will provide a fair test of the drug. As Eric Wieschaus [1] describes for the laboratory, "The wonderful thing is I know I'm back to normal because none of my experiments are working...it's back to the day-to-day struggle of getting it to work". In this regard, we have been interested to develop methods that could be used to

rebecker2008@comcast.net, greign@grc.nia.nih.gov

• Conflicts of Interest: REB is owner of Aristea Translational Medicine Corp. NHG has none.

preempt unnecessary failures of a clinical trial. With assurances that clinical trials “are working” as experiments, AD drug developers could, with greater confidence, then provide definitive tests of AD drug candidates.

Individual authors and international task force groups have analyzed (AD) clinical trials to identify reasons for repeated failures. Candidate causes have ranged widely: inadequate study population sizes, absence of biomarkers for the pathology, inadequate outcome measures and endpoint selections, ineffective drugs, failed target engagement, flawed study designs, inadequate Phase II studies, inclusion of subjects not truly AD pathology, and study of a disease stage too late for effective intervention [2-5]. In response to these issues, a task force recently proposed a shift in design of clinical trials to evaluate the efficacy of AD therapy candidates with early interventions among subjects both pre-dementia and pre-mild cognitive impairment [6]. In addition authors have called for other innovations in drug development methods [7-12]. Paul [7] proposed a shift of Proof of Concept into Phase I to better focus developments. Linsbury [8] integrated new drug library screening for target molecules with traditional cell culture and animal model testing to take full advantage of current resources. Klein [9] recently reemphasized opening Phase II to serendipity to overcome problems of not adequately grounding Phase III clinical trials. Pangalos [10] and Markou [11] separately called for increased validation of preclinical models as predictors for clinical applications. Collins [12], concerned with improved translations of laboratory discoveries into human benefits, recommended a shift to “targets ...relevant to multiple diseases” instead of current concerns more focused on a drug’s potential activity in one disease of interest.

We are concerned whether adequately full consideration has been given to characterization of the problem or problems that could account for the multitude of AD drug development failures in the last three decades [13] and to the impacts these problems might have on evaluations of methodological innovations. Aisen’s [6] proposal for earlier testing in response to his concerns that drugs may be effective for AD only if administered before the onset of clinical disease may be practical only if long term preventive trials can overcome any methodological weaknesses that have accounted for AD drug candidate failures. Even though Pangalos [10] and Markou [11] called for validation of preclinical tests as predictors for clinical phases of developments, they agree that, other than for validation as predictors of the biochemical effects of drugs, validation of preclinical predictors can occur only when an AD drug realizes efficacy in Phase III. Collins’ [12] focus on “targets ...relevant to multiple diseases” may not be immediately informative for the amyloid precursor protein (APP) metabolite, amyloid- $\beta_{(1-42)}$ ($A\beta_{42}$), related AD research since no known pathological mechanism seems able to account for the appearance of $A\beta_{42}$ related pathologies in the various diseases with increases in $A\beta_{42}$ concentrations in tissues: AD; Macular Degeneration; Cerebral Amyloid Angiopathy associated with Vascular Dementia; or post mild Traumatic Brain Injury late onset Chronic Traumatic Encephalopathic Dementia.

AD drug development may, for proper disposal of some problems able to hinder further advances towards a therapy, be facing a chicken or egg—which comes first—quagmire. Just as Pangalos and Markou recognize the need for a successful AD drug development to validate a preclinical test as a predictor of success, so too the confirmation in application of a

new method needed to overcome one or more barriers to the success of an AD drug will depend on successfully overcoming other barriers able to cause a failure of the test of the new method. In hopes of making a useful contribution to the conditions needed to insure the fairness of a test in AD of either a new drug or a new therapy, we used data from the failed development of phenserine to determine whether retrospective analyses could offer guidance towards distinguishing a clinical trial that failed due to an ineffective drug from a clinical trial that failed due to flaws in methods or conditions of the study. We regard this distinction as potentially useful in future evaluations of AD drug candidates and of innovations in methods for AD drug evaluations. Methods to validate that clinical trial conditions have provided fair and unbiased testing of the drug are needed to avoid AD drug candidates being inappropriately abandoned as failures. We hope to indicate some techniques that may be useful in the day-to-day process of getting AD drug developments to work [1].

Background

Phenserine

Phenserine, (-)- chirally pure, lipophilic, and orally bio-available, readily enters brain and reversibly inhibits acetylcholinesterase (AChE) [14]. Phenserine also reduces APP and A β ₄₂ levels by post-transcriptionally regulating APP mRNA translation at the 5'-untranslated region (5'-UTR) [15-17]. Reductions of A β ₄₂ up to 50% in cell culture [18], mice [19] and 20% to 37% in humans [20,21] were observed. The cytokines IL-1 and TGF β , and iron, each elevated in AD, up regulate APP via the 5'-UTR of APP mRNA [15-17]. In 32 healthy elderly volunteers, single doses of 5, 10, and 20 mg phenserine produced no serious adverse events, other than nausea and vomiting at the 20 mg dose [14]. In clinical trials investigators safely administered phenserine to over 500 humans for up to 18 months. Preclinical studies of phenserine provided strong expectations for success; yet, during Phase III clinical trial testing it was abandoned because clinical trials failed to evidence efficacy.

Phenserine was patented by the United States Government as an AChE inhibitor (AChE-I) for use in AD between 1995 and 1998. Then in 2000 an additional patent to cover the inhibition of amyloid precursor protein synthesis was added as phenserine entered human Phase I testing. After licensing for commercial development, between 2000 and 2006 the development proceeded into Phase III. With the failure to demonstrate efficacy in an analysis of the '06' phenserine outcomes—a 5-15 mg b.i.d., double-blind, placebo-controlled, n = 377, 26 week clinical trial conducted in Eastern Europe, Austria, Spain, and UK—the development of phenserine was abandoned. Subsequently the development rights to phenserine became, through acquisitions and purchases, the property of three different pharmaceutical firms each of which declined to pursue further development. In the published literature and in documentation of phenserine development we obtained from its final owner, we could not find evidence to distinguish whether phenserine was abandoned and remained abandoned because after failed clinical trials it was regarded as an ineffective drug, as a casualty of flawed development that could not be remedied by re-development, or for other reasons.

Methods

Materials

We searched Medline, the Internet, the records in our possession that were represented by the earlier acquiring companies as well as all the remaining phenserine development documentation, and the bibliographies of documents we uncovered to identify clinical phase phenserine research studies. We identified 20 potentially relevant publications on phenserine and 19 clinical study related planning and reporting documents on development. We then identified 3 ‘clinical’ phase publications—Phase I [14]; Kadir et al. [22]; and Winblad et al. [23]—and 12 development documents among which were documents containing full reports on the preclinical and clinical phenserine studies reported to the Food and Drug Administration. Of relevance to our interests were an analysis of the Phase III 06 Study including the performances at research sites in the 06 Study, a retrospective Status Report for 06/10 Studies focused on the implications of the development for redevelopment, and submissions to the FDA. Detailed analyses other than brief reports of failure to detect efficacy in 09 and 10 studies, abandoned after the 06 analysis, were not available. Since our interest was to analyze the clinical development for evidence to distinguish whether phenserine was abandoned as a failed drug or a victim of a flawed development, we chose to restrict our materials for study to the ‘clinically relevant documentations’ in our hands: the Investigator’s Brochure; Phase II Report to FDA; the final summary development analysis titled Status Report for 06/10 Studies; the 06 Study analysis including the performances at research sites in the 06 Study, the Kadir study protocol [22] which we kindly received from the principal investigator of the study Agneta Nordberg, the report of phase I by Greig [14] and a report of a final ‘close-out’ study organized by Winblad et al. [23] to document effects in a group of subjects continued in the double-blind after the decision to abandon further development.

Selection of documentation for further study

To familiarize ourselves with the overall development and possible causes for the abandonment decision each author separately read each of the ‘clinical’ publications and the ‘clinically relevant documentations’ materials. After independently concluding that the abandonment was not documented for cause in these available materials the authors also concurred that the analyses for cause would reflect the total available data by using the data and analyses contained in the Investigator’s Brochure; Phase II Report to FDA; the 06 Study analysis including the performances at research sites in the 06 Study, a final summary development analysis titled Status Report for 06/10 Studies; the Kadir study [22] and protocol and Winblad et al. [23].

Analyses

Based on earlier investigations into possible error sources able to interfere with AD drug developments [24] and additional risks identified more recently by individuals and task force reviews of AD drug developments [2-6], we undertook three approaches to the data. First we sought to relate the frequency of risks from inattention to error sources by phenserine investigators to the frequencies in other AD drug developments. We had earlier identified, from a study of 40 randomly chosen AD drug developments the frequency of consideration

by investigators that possible bias could have been introduced by error risks previously identified by others' and by us [24]. To establish possible similarities to these prior AD drug developments, using the methods reported in [24] we reviewed and evaluated for error risks the phenserine 'clinical' publications and the 'clinically relevant documentations' materials.

Second, given the viewpoint recently expressed by individual investigators and task forces—that AD drug developments have proceeded out from preclinical to clinical [10,11] and from Phase II to Phase III with insufficient preparations [3]—we sought to determine whether the phenserine developers could be considered to be adequately prepared to progress through the stages of development into Phase III. Using the methods in [24], but this time supplemented as necessary with concerns more recently raised—inadequate study population sizes, absence of biomarkers for the pathology, inadequate outcome measures and endpoint selections, ineffective drugs, failed target engagement, flawed study designs, inadequate Phase II studies, inclusion of subjects not truly AD pathology, and study of a disease stage too late for effective intervention [2-6]—we once again reviewed the 'clinical' publications and the 'clinically relevant documentations' materials to determine the specific phase, preclinical to phase III, of development in which an identified error risk occurred. For these first and second assessments each author separately pinpointed errors or error vulnerabilities. The authors then reconciled differences by consensus.

Finally, to determine whether errors in methods or procedures could account for the outcome from the unpublished Phase III 06 study that had provided grounds for the investigators to abandon the phenserine development, we analyzed the data from research sites for possible effects on the major components, variance and accuracy, determining the p value associated with the outcome. To proceed with these analyses we first inspected the data with an exploratory data analysis, as recommended by Tukey [25], to determine the structural sources of excessive variance, inaccuracy, or bias affecting these factors. We then, using Fisher Exact Statistics [26], examined each of these factors by research sites to determine whether statistically significant, $p = 0.05$ or less, relationships occurred. These analyses tested the hypothesis that methods or conditions had undermined the fairness of the study as an unbiased test of the research hypothesis and that the outcome was an effect from the methods or conditions and not the proposed independent variable phenserine.

Results

Does the phenserine development share flaws characteristic among other AD drug developments?

We found little evidence to support the view that investigators regarded error risks as issues to be addressed preemptively or in interpreting study outcomes (Table 1).

Did the phenserine developers proceed onto subsequent Phases of development adequately prepared to insure, as best as possible, success?

We found in each stage identified errors that remained unaddressed as the developers moved on to the subsequent stage of development (Table 2).

Can methodological lapses in the phenserine development account for the failed studies?

In our initial exploratory data analysis we found an Alzheimer's Disease Assessment Scale-Cognitive Subscale (ADAS-Cog) [27] variance of ± 4.5 points or greater in each of the phenserine clinical trials including the 06 trials. This interested us since we had earlier identified this degree of variance as potentially problematic [28], found that control of this degree of variance associated was associated with success in a comparison of four trials [29], and had noted that the recent Huperzine A trial [30] had projected its power calculations for the ADAS-Cog based on 4.5 points standard deviation in an earlier study [31]. We earlier questioned whether adjustments to N, suggested by Cohen as necessary but not desirable to achieve power [32], were causing investigators to overlook compromises to validity from excessive variance. To test for variance or rater imprecision effects in the 06 clinical trial we compared the variance in placebo groups at research sites to predictions from our previously estimated Standard Error of Measurement (SEM) for the ADAS-Cog [27]. For this comparison we adjusted the SEM using group sizes to calculate mean SEMs for the placebo groups at the research sites. This identified 5 sites where the mean placebo ADAS-Cog ratings exceeded the mean SEM adjusted for group size and 7 sites where they did not. We then compared these site groups for possible effects on the outcome.

The exploration of data also revealed that in each of the five phenserine clinical trials for which we had data the placebo groups improved over baseline [22,23]. Placebo groups generally should decline over time since AD is a progressive neurodegenerative disorder. Since placebo group improvements have been identified as suspect [33], but not proven, risks to validity we compared those sites with placebo group improvements to those without placebo groups improvements for systematic effects on the outcome. We also had found that some placebo group subjects improved over their baseline scores at the first double-blind post-baseline clinical status rating. Since these improvements are consistent with raters inflating pathology to qualify subjects for entry or subjects who are not stabilized to study conditions we analyzed this occurrence for a relationship to outcomes.

A non-statistically significant relationship occurred between excess variance and placebo group improvement (Table 3i). Statistically significant relationships occurred between excess variance and failures of active drug treated subjects to improve on outcome ratings (Table 3ii), between placebo group responses and study outcomes (Table 3iii), and between placebo group initial post-baseline and subsequent placebo group double-blind ratings (Table 3iiii).

Discussion and Conclusions

Does the phenserine development share flaws characteristic among other AD drug developments?

Given the similar low levels of attention to published error risks in both our earlier study review [24] and the phenserine development we conclude that phenserine could be considered to be at similar risks for invalidation by error as were the 40 other AD drug developments we had studied. Recent critiques of selected AD studies have presented

evidence to support that flawed methods were responsible for outcomes, a risk that phenserine could share.

Did the phenserine developers proceed onto subsequent Phases of development adequately prepared to insure, as best as possible, success?

We found in each stage errors and vulnerabilities to errors that remained unaddressed at the time that the developers moved on to the subsequent stage of development. Similar to our reasoning about errors, we interpret this finding as supporting the recent criticisms by investigators and Task Force members of Preclinical and Phase II preparations for AD clinical trials.

Can methodological lapses in the phenserine development account for the failed 06 clinical trial?

We documented statistically significant effects from methods or conditions on the 06 clinical trial outcomes. The Table 3iiii analysis does not identify the source for the placebo group improvements under blind rating conditions. Preemptively, we have shown that extended baseline evaluations can accommodate subjects to study conditions and provide an unbiased estimator for each individual subject's initial state thus increasing power [28,29]. This baseline would not necessarily identify deliberate rating inflation to qualify subjects for a study.

These data and their statistically significant relations to outcomes at research sites render this Phase III clinical trial invalid as a test of the independent variable, the drug, and potentially put under suspicion of being invalid other AD drug development studies we found equally vulnerable to error risks [24]. It is possible that AD investigators currently give too little attention to mistakes and biases that can invalidate their efforts.

We conclude that phenserine was abandoned not because the drug failed to exhibit efficacy but because the methodology failed to provide a fair and unbiased opportunity for the drug to express efficacy or lack of efficacy. Given the similarities to other AD drug developments, some abandoned AD drug developments may have resulted from invalidation of the studies rather than fair tests for efficacy

Are commonly accepted standards for variance and placebo group responses to be questioned?

Our data are consistent with a concern that occurrences in AD studies of the degrees of ADAS-Cog variance and placebo group improvements that invalidated the phenserine trial may have more widespread implications for the validity of clinical trials. We encourage further studies by clinical neuroscience of these and other factors to determine whether Cohen [32] and others since have too uncritically taken placebo group improvements and variance as benign: controlled by randomization or safely "tucked" away by increasing N. As we earlier implemented in one of our AD drug developments [34,35], steps to control variance may be of equal importance with steps shown necessary to control inaccuracies [36] if investigators intend to provide themselves and others confidence in outcome analyses.

A way forward

Overconfidence that observations can speak for themselves may have allowed the phenserine investigators to move too directly to test for phenserine efficacy and then, without testing for validity, to abandon the development. This direct approach to efficacy was described in 2007 as the basic guidance available to AD drug development investigators: “Neil Buckholtz, Chief of the Dementias of Aging Branch, National Institute on Aging, NIH (Bethesda, MD), said the pharmaceutical industry had little choice. “This is basically a ‘throw the spaghetti against the wall’ strategy.... We just have to try these various approaches. It's very time consuming, very expensive, but it's the only way we'll know if things work or not.”” [37]. We have demonstrated methods by which the validity of clinical trials can be tested. We propose that such testing could avoid the abandonment of drugs inappropriately.

In contrast with this expression of confidence in the strength of current AD drug development methods to provide unbiased estimates of efficacy, our analyses indicate that efficacy evolves more complexly out from a more successful completion of stepwise, often iterative, two steps forward and one step back, multifactorial and prolonged, scientifically restrained inquiries. As physicists have found, confidence in ascribing the outcome to the independent variable arises only after possible sources of error in methods and conditions of the experiment have been excluded [38]. We find support for the need for investigators continually to plan, monitor, revisit, and reconfirm the grounding of each new step taken towards the efficacy goal. Since the success of each later stage of a drug development depends on the outcomes at earlier stages, risks from biases will cumulatively increase over time. As we have suggested in earlier publications, AD drug developers could emulate the nuclear and aviation industries and give increased attention to preemptions, monitoring, and corrective interventions aimed to control risks to validity [39]. It may be timely to update Jacob Cohen, father of the *d* effect size estimate. “We...,” *in view of the error risks to our studies*, “should be far more concerned with why and what and how well we are measuring, manipulating conditions, and selecting our samples” *and otherwise conducting our studies* [32].

Acknowledgments

NHG is supported by the Intramural Research Program, National Institute on Aging, NIH. REB is the founder of Aristea Translational Medicine Corp., which has ownership of the phenserine documentation, and received no funding support for this work.

References

1. Kreeger, KY. [Viewed August 15, 2012] Overdue, first-time recognitions mark '95 nobel science, peace prizes. The Scientist. Nov 13. 1995 <http://classic.the-scientist.com/?articles.view/articleNo/17647/>
2. Khachaturian ZS, Barnes D, Einstein R, Johnson S, Lee V, Roses A, et al. Developing a national strategy to prevent dementia: Leon Thal Symposium 2009. *Alzheimers Dement*. 2010; 6:89–97. [PubMed: 20298968]
3. Vellas B, Aisen PS, Sampaio C, Carrillo M, Scheltens P, Scherrer B, et al. Prevention trials in Alzheimer's disease: an EU-US task force report. *Prog Neurobiol*. 2011; 95:594–600. [PubMed: 21925234]

4. Vellas B, Hampel H, Rougé-Bugat ME, Grundman M, Andrieu S, Abu-Shakra S, et al. Alzheimer's Disease Therapeutic Trials: EU/US Task Force Report on Recruitment, Retention, and Methodology. *J Nutr Health Aging*. 2012; 16:339–45. [PubMed: 22499454]
5. Hampel H, Wilcock G, Andrieu S, Aisen P, Blennow K, Broich K, et al. Biomarkers for Alzheimer's disease therapeutic trials. *Prog Neurobiol*. 2011; 95:579–93. [PubMed: 21130138]
6. Aisen PS, Andrieu S, Sampaio C, Carrillo M, Khachaturian ZS, Dubois B, et al. Report of the task force on designing clinical trials in early (predementia) AD. *Neurology*. 2011; 76:280–6. [PubMed: 21178097]
7. Paul SM, Mytelka DS, Dunwiddle CT, Persinger CC, Numos BH, Lindborg SR, et al. How to improve R&D productivity: the pharmaceutical industry's grand challenge. *Nature Reviews: Drug Discovery*. 2010; 5:203–14.
8. Lansbury PT. Back to the future: the 'old fashioned' way to new medications for neurodegenerations. *Nature Rev Neurosci*. 2004; 5:S51–S57.
9. Klein DF. The loss of serendipity in psychopharmacology. *JAMA*. 2008; 299:1063–5. [PubMed: 18319418]
10. Pangalos MN, Schechter LE, Hurko O. Drug development for CNS disorders: Strategies for balancing risk and reducing attrition. *Nat. Rev. Drug Discov*. 2007; 6:521–32. [PubMed: 17599084]
11. Markou A, Chiamulera C, Geyer MA, Tricklebank M, Steckler T. Removing obstacles in neuroscience drug discovery: The future path for animal models. *Neuropsychopharmacol*. 2009; 34:74–89.
12. Collins FS. Reengineering translational science: the time is right. *Sci Transl Med*. 2011; 3 90cm17.
13. Lindner, MD.; McArthur, RA.; Deadwyler, SA.; Hampson, RE.; Tariot, PN. Development, optimization and use of preclinical behavioral models to maximize the productivity of drug discovery for Alzheimer's disease. In: McArthur, RA.; Borsini, F., editors. *Animal and Translational Models for CNS Drug Discovery: Neurologic Disorders*. Academic Press; San Diego: 2008. p. 93-157.
14. Greig NH, Ruckle J, Comer P, Brownell L, Holloway HW, Flanagan DR, et al. Anticholinesterase and pharmacokinetic profile of phenserine in healthy elderly human subjects. *Curr. Alzheimer Res*. 2005; 2:483–92. [PubMed: 16248851]
15. Shaw KT, Utsuki T, Rogers J, Yu QS, Sambamurti K, Brossi A, et al. Phenserine regulates translation of beta-amyloid precursor protein mRNA by a putative interleukin-1 responsive element, a target for drug development. *PNAS*. 2001; 98:7605–10. [PubMed: 11404470]
16. Lahiri DK, Chen D, Vivien D, Ge YW, Greig NH, Rogers JT. Role of cytokines in the gene expression of amyloid beta-protein precursor: identification of a 5'-UTR-binding nuclear factor and its implications in Alzheimer's disease. *J Alzheimers Dis*. 2003; 5:81–90. [PubMed: 12719626]
17. Rogers JT, Leiter LM, McPhee J, Cahill CM, Zhan SS, Potter H, et al. Translation of the Alzheimer amyloid precursor protein mRNA is up-regulated by interleukin-1 through 5'-untranslated region sequences. *J Biol Chem*. 1999; 274:6421–31. [PubMed: 10037734]
18. Mikkilineni S, Cantuti-Castelvetri I, Cahill CM, Greig NH, Rogers JT. The anti-cholinesterase phenserine and its enantiomer posiphen as 5' untranslated region directed translation blockers of the Parkinson's alpha synuclein expression. *Parkinson's Disease*. 2012 In press.
19. Venti A, Giordano T, Eder P, Bush AI, Lahiri DK, Greig NH, et al. The Integrated Role of Desferrioxamine and Phenserine Targeted to an Iron-Responsive Element in the APP-mRNA 5'-Untranslated Region. *Ann. N.Y. Acad. Sci*. 2004; 1035:34–48. [PubMed: 15681799]
20. Sambamurti, K.; Prakasam, A.; Anitha, S.; Venugopa, C.; Cullen, E.; Zhou, Y., et al. Oral phenserine tartrate treatment is associated with reduced plasma Ab 1-42 in a phase I clinical study. In: Khalid, IK.; Winblad, B.; Avila, J., editors. *Alzheimer's Disease New Advances*. Medimond, Via Maserati 5; Bologna, Italy: 2007. p. 571-5.
21. Thatte U. Phenserine Axonyx. *Current Opinion in Investigational Drugs*. 2005; 6:729–39. [PubMed: 16044670]

22. 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. [PubMed: 18300284]
23. Winblad B, Giacobini E, Frölich L, Friedhoff LT, Bruinsma G, Becker RE, et al. Phenserine efficacy in Alzheimer's disease. *J Alzheimer Dis*. 2010; 22:1201–8.
24. Becker RE, Greig NH, Giacobini E. Why Do So Many Drugs for Alzheimer's Disease Fail in Development ? Time for New Methods and New Practices? *J Alzheimer Dis*. 2008; 15:303–25.
25. Hoaglin, DC.; Mosteller, F.; Tukey, JW. *Understanding Robust and Exploratory Data Analyses*. John Wiley & Sons; New York, NY: 2000.
26. Siegel, S. *Non-Parametric Statistics*. McGraw Hill; New York, NY: 1956.
27. Rosen WG, Mohs RC, Davis KL. A new rating scale for Alzheimer's disease. *Am J Psychiatry*. 1984; 141:1356–64. [PubMed: 6496779]
28. Becker RE, Markwell S. Problems arising from the generalizing of treatment efficacy from clinical trials in Alzheimer's disease. *Clin Drug Invest*. 2000; 19:33–41.
29. Becker RE. Lessons from Darwin: 21st Century designs for clinical trials. *Curr Alzheimer Res*. 2007; 4:458–67. [PubMed: 17908050]
30. Rafii MS, Walsh S, Little JT, Behan K, Reynolds B, Ward C, Jin S, et al. A phase II trial of huperzine A in mild to moderate Alzheimer disease. *Neurology*. 2011; 76:1389–94. [PubMed: 21502597]
31. Aisen PS, Schneider LS, Sano M, Diaz-Arrastia R, van Dyck CH, Weiner MF, et al. High-dose B vitamin supplementation and cognitive decline in Alzheimer disease: a randomized controlled trial. *JAMA*. 2008; 300:1774–83. [PubMed: 18854539]
32. Cohen J. Things I have learned (so far). *Amer. Psychologist*. 1990; 45:1304–12.
33. Schneider LS, Sano M. Current Alzheimer's disease clinical trials: methods and placebo outcomes. *Alzheimers Dement*. 2009; 5:388–97. [PubMed: 19751918]
34. Becker RE, Colliver JA, Markwell SJ, Moriearty PL, Unni LK, Vicari S. A double-blind, placebo-controlled study of metrifonate, an acetylcholinesterase inhibitor, for Alzheimer's disease. *Alzheimer Dis Assoc Disord*. 1996; 10:124–31. [PubMed: 8876775]
35. Becker RE, Colliver JA, Markwell SJ, Moriearty P, Unni LK, Varney A, et al. A six month study of the effects of metrifonate on cognitive decline in Alzheimer's disease. *Alzheimer Dis Assoc Disord*. 1998; 12:54–7. [PubMed: 9539412]
36. Engelhardt N, Feiger AD, Cogger KO, Sikich D, DeBrotta DJ, Lipsitz JD, et al. Rating the raters: assessing the quality of Hamilton rating scale for depression clinical interviews in two industry sponsored clinical drug trials. *J Clin Psychopharmacol*. 2006; 26:71–4. [PubMed: 16415710]
37. McDermott, T. [Viewed August 15 (2012)] Scientists can't get their minds around Alzheimer's. *Los Angeles Times*. Dec 27. 2007 <http://www.latimes.com/news/printedition/front/la-na-alzheimers27dec27,0,1738791.story>
38. Duhem, P. *The Aim and Structure of Physical Science*. Princeton University Press; Princeton: 1991.
39. Becker RE, Greig NH. Why so few drugs for Alzheimer's disease? Are methods failing drugs? *Curr Alzheimer Res*. 2010; 7:642–51. [PubMed: 20704560]

Table 1

Evidence for investigator vigilance against errors

Potential Risk Categories	Phenersine Drug Development % compliant	40 Randomly Chosen AD Drug Developments [24] % compliant
1. Drug, mechanisms, pharmacological activities, design, and publications	32	41
2. Dosing	0	16
3. Research subjects	15	19
4. Outcome measures	18	10
5. Research sites and investigators	8	17
6. Protocol controls of methods and clinical transfer	10	3

Table 2

Preparedness for moving to the next stage of the phenserine development

Phase Interface/ Unresolved Issues	Preclinical - Phase I	Phase I - Phase II	Phase II - Phase III
Aim for Development	AChE-I or APP-SI? No link of human dosing effects to disease modification aims	AChE-I or APP-SI? FDA request for disease modification Fast-Track Approval; however, Human APP-SI dosing study not initiated until Phase III studies in process	AChE-I or APP-SI?
Preclinical Studies	Failure to select lead compound		
Dosing	Dosing predictions for humans incomplete: no dosing model; criteria for accepting human target engagement not established.	Target Engagement not demonstrated for b.i.d. dosing.	Phase II studies not completed Dosing not established and confirmed. APP-SI dosing studies undertaken during Phase III after disease modification study initiated.
Outcome Measures	No animal model data for Paired Associates Learning	Positive PAL outcomes not confirmed	
Certainty of Interpretations		Not statistically significant outcomes	Outcomes not available when Phase III entered
Power	Inadequate projected power either due to excessive variance or insufficient N assumed.	Inadequate projected power either due to excessive variance or insufficient N	Inadequate projected power either due to excessive variance or insufficient N
Subjects		Possible rater bias qualifying subjects or subject attention bias	
Variance		Possibly excessive inter-site variance not addressed	Possibly excessive inter-site variance not addressed
Placebo Group Outcomes	Unexplained Placebo Group Improvement	Unexplained Placebo Group Improvement	Unexplained Placebo Group Improvement
Protocol	Inadequate multiple dose and drug target kinetics testing in relation to providing target engagement	Protocol not adequate to test for stated aim of disease course modification.	

AChE-I is acetylcholinesterase inhibitor

APP-SI is amyloid precursor protein synthesis inhibitor

b.i.d. is three times per day dosing

N is number of subjects

PAL is the Paired Associates Learning test

Table 3

Variance and Placebo Group Analyses

3i. Rater variance and placebo group ratings at Study 06 sites			
For 11 sites (2 of 11 composed of multiple smaller sites)	Variance of Placebo Ratings Controlled (Mean Ratings within 95% CI for Mean SEM)	Variance of Placebo Ratings Uncontrolled (Mean Ratings Exceed 95% CI for Mean SEM)	Sums
Placebo Group Worsened	5	2	7
Placebo Group Improved	1	3	4
Sums	6	5	F.T.E. * $p = n.s.$

3ii. Rater variance and subject's active drug responses at Study 06 sites.			
For 11 sites (2 of 11 composed of multiple smaller sites, 2 drug groups at each site- 10 and 15 mg. b.i.d.)	Variance of Placebo Ratings Controlled (Mean Ratings within 95% CI for Mean SEM)	Variance of Placebo Ratings Uncontrolled (Mean Ratings Exceed 95% CI for Mean SEM)	Sums
Active Drug Group Improved over Placebo Group	9	3	12
Active Drug Group the Same as or Worse than Placebo Group	3	7	10
Sums	12	10	F.T.E. * $p = 0.05$

3iii. Placebo group ratings and subjects' active drug responses at Study 06 sites.			
For 11 sites (2 of 11 composed of multiple smaller sites, 2 drug groups at each site- 10 and 15 mg. b.i.d.)	Placebo Group Worsened	Placebo Group Improved	Sums
Active Drug Group Improved over Placebo Group	11	1	12
Active Drug Group the Same as or Worse than Placebo Group	3	7	10
Sums	14	8	F.T.E. * $p = 0.005$

3iiii. Placebo group rating relations to baseline ratings			
	First Post-Baseline Placebo Rating Equal or Less than Baseline	First Post-Baseline Placebo Rating Improved over Baseline	Sums
Placebo Groups Improved during Double-Blind	0	4	4
Placebo Groups Declined during Double-Blind	6	1	7

3iiii. Placebo group rating relations to baseline ratings			
	First Post-Baseline Placebo Rating Equal or Less than Baseline	First Post-Baseline Placebo Rating Improved over Baseline	Sums
Sums	6	5	F.T.E. * $p = 0.05$

* F.E.T. is Fisher's Exact Test ([26] pp. 257; 263; 266; 256)

SEM is the Standard Error of Measurement adjusted for the group sizes