Published in final edited form as:

Curr Alzheimer Res. 2010 November 1; 7(7): 642–651.

Why So Few Drugs for Alzheimer's Disease? Are Methods Failing Drugs?

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

¹Aristea Translational Medicine Corporation, Freeport, ME 04078; USA

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

Abstract

Recent studies of Alzheimer's disease (AD) and other neuropsychiatric drug developments raise questions whether failures of some drugs occur due to flaws in methods. In three case studies of recent AD drug development failures with phenserine, metrifonate, and tarenflurbil we identified methodological lapses able to account for the failures. Errors in complex systems such as drug developments are both almost inescapable due to human mistakes and most frequently hidden at the time of occurrence and thereafter. We propose preemptive error management as a preventive strategy to exclude or control error intrusions into neuropsychiatric drug developments. We illustrate the functions we anticipate for a preemptive error management preventive strategy with a checklist and identify the limitations of this aspect of the proposal with three drug examples. This strategy applies core scientific practices to insure the quality of data within the current context of AD drug development practices.

Keywords

Alzheimer's disease; Alzheimer clinical trial; error management; checklist; tarenflurbil; metrifonate; phenserine

INTRODUCTION

Setting

Recent late-stage failures of drugs regarded as promising for Alzheimer's disease (AD)—tramiprosate [1-3], phenserine [4,5], tarenflurbil [6,7], dimebon [8,9]—continue the disappointments in neurological and psychiatric or neuropsychiatric clinical trials (CT) and drug developments over the last two decades. In 2008 two separate groups identified, after reviews of attempts to develop a drug for AD, over 100 [10] and 172 [11] drug development failures. In this paper we take the evidence from continued failures of neurological and psychiatric drug candidates, compounds regarded originally as promising by industry and leaders in drug developments, as potentially indicative of problems able to block therapeutic progress and advancement of scientific knowledge of the mechanisms open to therapeutic interventions beneficial for AD [12].

© 2010 Bentham Science Publishers Ltd.

*Address correspondence to this author at the Aristea Translational Medicine Corporation, Freeport, ME 04078; USA; rebecker2008@comcast.net.

Competing interests: Robert Becker has an equity interest in Aristea Translational Medicine Corporation that was formed as a platform for clinically investigating drugs.

Successful drug developments occur as time consuming results from complexly managed, longitudinally interdependent, cumulatively informative, diverse practices completed in specific sequences. In both simple tasks and complex enterprises such as drug developments, a failure almost instinctively evokes for the persons involved considerations of whether the method or methods used are responsible for the failure. In the dark, unable to open a lock with a key selected from a key ring, the homeowner checks whether a methodological error, such as selecting the wrong key, has been committed. In complex processes similar in structure to drug developments, such as operations of nuclear plants or aircraft, experience shows that human lapses, mistakes, slips and other deficits in planning and execution readily leave the systems vulnerable to errors. In complex processes, errors characteristically remain undetected by operators or by a system not armed specifically to be vigilant against operator errors. Consequently, complex operating systems in nuclear plants and aircraft, to operate safely and effectively, require special planning to provide error-suppressing supports for operators [13].

In this paper we consider the possibility that human errors may contribute to the repeated failures of AD drug developments. We apply experiences from the nuclear and aircraft industries to characterize error vulnerability for neuropsychiatric drug developments. From these analyses we propose some simple steps for structuring an approach to controlling error sources that have been identified in recent investigation reports as possible sources for drug failures [10,12,14]. These proposed methods adapt for neuropsychiatric drug developments the strategies found useful for controlling methodological errors in other complex undertakings; particularly, the revision of methods user-hostile to make them user-friendly [13, pp. 234-237]. User hostile methods, found commonly in drug developments, allow users to commit mistakes that will become expressed as undetected errors in data sets or practices. To overcome these dispositions towards error effects resulting from inevitable human mistakes, the nuclear and aircraft industries provide user-friendly supports to operators. User-friendly methods warn operators that a choice risks introducing an error or, when possible, prevent operators from committing errors. If methods in AD drug developments undermine the independent variable, the drug under investigation, expressing its effects without bias from the conditions of testing, then some current neuropsychiatric clinical trial (CT) outcomes are potentially invalid and neuropsychiatry's CT evidence for efficacy or lack of efficacy may be misleading for investigators and practitioners. Methodological errors or vulnerabilities to errors are not without consequences for neurology, psychiatry, and the basic pharmacological sciences that depend on the successes of clinical studies to confirm the relevance of preclinical methods.

BACKGROUND

In a 2008 investigation Becker, Greig, and Giacobini [10] selected randomly for critical review from among protocols or reports on over 100 AD drug development attempts: 10 each of studies associated with an approved, failed, or in development drug for AD, and 10 studies in mild cognitive impairment. Each of the selected study documents was then screened for statements addressing one of 56 potential error sources we or others had identified as part of earlier investigations into neuropsychiatric drug development failures [10]. Table 1 shows representative items from the survey to illustrate sources of error assessed in these document reviews and the typical published rates of evidence that the issue had been or was being planned in protocols to be addressed specifically by investigators. For almost all risks less than 10-20% of documents indicated that investigators had given or would give specific consideration to each error source and its possible effects on the validities of their studies.

Three neuropsychiatric drug developments suggest that this inattention to problems of error intrusions may not be benign in effects on validity of conclusions available to investigators (Table 2). For the two of these three drugs that failed in Phase III AD CTs—phenserine and metrifonate—other small CTs evidenced efficacy [15-17; Unpublished data]. Studies of methods used for each of these drugs [5,10,12,17] suggested problems with variance and dosing sufficient to impugn as possible sources for the developmental failures methodological misjudgments of investigators and errors introduced by these misjudgments. The case study of tarenflurbil challenges the investigators' conclusions that the drug failed in its Phase III CT and suggests instead that, because of methodological weaknesses, no firm conclusions on efficacy are appropriately reached [12]. Similar concerns have been raised retrospectively by principal investigators for other multisite AD CTs failed because of the drug being abandoned by the commercial sponsor in spite of evidence suggestive of possible drug benefits [2,18].

This case-study evidence suggests to us that, in addition to these specifically considered AD drug developments, not presently appreciated problems with methods may underlie other recent AD drug development failures. These methods known to be vulnerable to error intrusions are being used in currently ongoing drug developments as documented by the numbers of developments and CTs we have cited as sources for evidence of errors and their effects. Without actions to overcome these vulnerabilities to errors, currently active and future AD drug developments may be at unnecessary risks from failures due to unaddressed errors. The problems include:

- 1) That drug development methods are too often “user-hostile” and invite undetectable user mistakes to introduce errors into the drug development,
- 2) That these sources allowing errors to intrude are not evident at the time they open drug developments to errors,
- 3) That, since intrusions of errors into drug developments cannot be detected and corrected while a drug development is in process or when outcomes are interpreted, as James Reasons concludes in *Human Error* [13, p. 250], this “insidious concatenation of latent human failings that are an inevitable part of any large organization” results in individual and combined errors able to invalidate AD studies, CTs, and drug developments, and finally,
- 4) That, since methodological errors are possible sources for experimental outcomes in drug developments, the design-related chances of error add to statistical estimates of Type I and II measurement-related errors making the certainty of conclusions from AD CTs always less than the p value adopted in the statistical model.

Table 3 provides some specific examples of how errors characteristically express themselves in drug developments and the consequential need for preemptive interventions.

Basic to the understanding of the model we use for how errors arise in drug developments is James Reason's model [13, p. 206] analogous with what would be expected with a light shined on slices of Swiss Cheese Fig. (1). Reason developed his model using studies of why mistakes occur for human operators and how these mistakes cause or facilitate disasters in nuclear power plants or aviation. The model identifies as crucial targets for interventions the latent deficits in defenses against mistakes that express themselves as errors able to affect the operating system. Predictions from Reason's model, from empirical studies, and from considerations of everyday personal experiences coincide in conclusions that errors will be almost inevitable in any system complex enough to over-tax or even stress the mental resources of humans who interact with the processes. Responses designed into systems and choices taken by even the best trained and experienced human participants, will not always,

as stresses accrue for humans from a complex system operating over time, be correct for the situation at hand. Either or both human participates, as operators or investigators, and the system itself will show limitations in planning that do not take into account a particular unanticipated circumstance. A human with responsibilities for outcomes will take an action not functional for the circumstances at hand. Given this expectation, confirmed by studies of nuclear power plant disasters and airplane crashes, in complex ongoing operating systems humans require “user-friendly” supports from the system [13, pp. 234-237]. User-friendly practices and procedures either inform the user that the action he or she is about to undertake will risk an error or they prevent the user from taking such an action. For AD drug developments and their component studies, including CTs, planning needs as much as possible to seek to include practices with error preemptive, user-friendly characteristics. This anticipatory planning asks planners and designers to prevent error intrusions by removing or controlling system vulnerabilities to errors being introduced by the almost inevitable human operator mistakes.

Drug developments are particularly vulnerable to errors because almost all the practices and methods available at all phases of drug developments provide little or no safeguards against investigators misusing procedures or introducing erroneous data. For example, AD rating scales allow any values in the range offered to be entered for a subject without regard to accuracy, precision, care on the part of the rater to comply with any protocols governing administration of the scale, willingness of the rater to provide falsified data, and so forth. For most errors identified as intruding into drug developments, we and others have found little suggestion that they can be detected and corrected at or after the time of their occurrence. Consequently, error management in neuropsychiatric drug developments must be preemptive. Successful preemption turns user-hostile sources for vulnerabilities into user-friendly indicators of risks or methods able to prevent latent vulnerabilities to errors being introduced into the drug development system's operations. A vulnerability to a mistake becoming an undetected error that is then able to invalidate a drug development or study is replaced whenever possible with a user-friendly procedure or task capable person able to prevent an investigator's mistake or the consequences from the mistake from occurring. Current understanding of errors in drug developments and their characteristics indicate to us that latent defects in defenses against error intrusions—practices and methods that are user-hostile or persons, by virtue of limitations in skills who are not research qualified—must be removed, controlled, or countered as best available in planning and early stages of drug developments. These early interventions prevent or reduce vulnerabilities of later stages of drug developments to undetected error intrusions.

A ROLE FOR CHECKLISTS?

Experiences in other error-sensitive industries indicate that AD investigators need to operate with awareness that in neuropsychiatric drug developments 1) errors will affect drug developments unless specifically addressed and 2) interventions can apply knowledge gained from earlier experiences and research into specific errors and the consequences from these errors for drug developments [19]. Checklists need to be tailored to raise as a topic for consideration each already described potential error and by error vulnerabilities specific to the particular conditions imposed by a drug being developed. So molded, checklists provide interested investigators a potentially useful tool to support error-preempting attitudes and intentions. As we model in our compiled error list [19], the problem posed by an error judged relevant by an investigator for a particular drug development can often be characterized using past error studies. This characterization helps to identify what vulnerability-target can be addressed with preemptive-preventive interventions. A checklist supports the user to focus attention on immediate issues—the opportunity to avoid latent vulnerabilities—to prevent later difficulties from error intrusions.

In our experience applications of checklists can be helpful both planning and reviewing progress at major milestones in drug developments. The reasoning structure available generically is illustrated for important representative issues in Table 4. For an issue of concern a user must implement a countermeasure to overcome ‘holes’ or lapses in defenses against this error source, estimate what probability for failure remains in spite of these countermeasures, and then decide whether this remaining risk is acceptable considering that this probability of failure will combine with other probabilities associated with latent failures to reduce certainty available for the overall drug development.

CHECKLIST REVIEWS OF THREE REPRESENTATIVE DRUG DEVELOPMENT FAILURES

Tarenfluribil

Tarenflurabil ((*R*)-flurbiprofen or (*R*)-2-(3-Fluoro-4-phenylphenyl)propanoic acid) is the chirally pure, single (*R*) enantiomer of the racemic (*R,S*), non steroidal anti-inflammatory agent, flurbiprofen. Tarenflurabil lacks COX-1 and COX-2 activity, is minimally converted to its (*S*) enantiomeric form *in vivo* and has been reported to possess γ -secretase modulating activity to lower amyloid- β peptide levels [6].

Myriad Corporation investigators had available to them Phase II CT data on which to base further steps in the development of this compound [6]. Although there are a range of concerns with the adequacy of the tarenfluribil Phase II CT as preparation for Phase III [7,12] the choice of dosing based on Phase II illustrates the use of a checklist item and its potential limitations. Let us assume a relatively high confidence in the dose selected and used by Myriad investigators for Phase III. The investigators effectively give a low *p* value to possibilities that dosing had not been adequately optimized to go forward and accepted this *p* value as a risk to validity. On the other hand, viewing the pharmacological data available to these investigators Fig. (2) the straight line of the ascending improvements in the outcome measure at the highest dose tested, expressed in Fig. (2) as the C_{max} , allow the possibility that more efficacy may occur at even higher doses than those within the tested range that were selected for the Phase II CT. These tested doses were described as safe and well tolerated in both AD and a prostate cancer trial [6, and phase 2a, trial no. EPR-7869-0003/0004 data on file, Myriad Pharmaceuticals], allowing pre-Phase III CT testing for dosing using further dose-escalations. Therefore, it does not seem unreasonable for a pharmacologist to estimate a *p*=0.50 that the selected dosing may not express the true efficacy available with this drug. The pharmacologist would be concerned that the efficacy indicated by Fig. (2) data may not be adequate to power a CT and the CT may fail because of this oversight. This associated risk estimate a pharmacologist might reasonably provide—that, in one chance out of two, a CT designed using these dosing assumptions could fail due to inadequate dosing—could be expected to suggest to Myriad investigators that at the conclusion of Phase II studies they are not prepared for Phase III. From this it can be seen that a checklist item, while useful as an indicator that an issue needs to be addressed, cannot overcome choices based on personal opinions, mistaken reasoning, empirically uninformed beliefs, or other than scientific grounding.

Metrifonate

Metrifonate (trichlorfon or (*R,S*)-2,2,2-trichloro-1-dimethoxyphosphoryl-ethanol) is an irreversible organo-phosphate acetylcholinesterase inhibitor that was effectively used in the treatment of schistosomiasis, a parasitic disease endemic in many tropical countries, and found effective following once weekly administration in AD [20-23].

Bayer Pharmaceutical investigators had available to them wide knowledge of the drug class organophosphates to which metrifonate belongs and publications relevant to development of a safe dosing schedule to allow testing of metrifonate for AD [20-23]. In our hands these sources provided us great certainty to expect from specific dosing schedules used with large populations some, if not many, cases of organophosphate toxicity of sufficient severity to preclude drug approval should efficacy for AD be shown. We chose for our CTs [15,16] weekly dosing Fig. (3) to allow recovery of enzyme activities we regarded as important to avoid organophosphate toxicity. For their development, Bayer chose daily dosing. As Fig. (3) documents the weekly dosing selected for the Becker *et al.* [15,16] CTs seemed both practical and able to avoid with an adequately safe margin of error degrees of enzyme inhibition associated with organo-phosphate toxicity. The weekly doses in these two small CTs showed evidence supporting efficacy against AD and safety.

One can assume that Bayer investigators chose a daily dosing regimen within the dose range shown in Fig. (3) with p estimates of risks due to dosing acceptable to them. The Bayer CTs, in spite of showing efficacy for metrifonate against AD, failed to win FDA NDA approval because of toxicity at higher doses. Checklists can raise issues to be addressed as potential sources for invalidation of drug developments but cannot exclude other influences such as business advantages to sales that could accrue with more frequent doses or other investigator bases for choices of methods.

Phenserine

Phenserine ((-)-phenserine or (-)-phenylcarbamoylseroline) is a non-competitive reversible acetylcholinesterase inhibitor that has independent post-transcriptional inhibitory effects on amyloid- β precursor protein synthesis, lowers amyloid- β generation [17], and was developed within the National Institute on Aging.

After licensing phenserine, Axonyx Corporation investigators chose to advance phenserine into an initial Phase III CT without determining the optimal dosing required to allow a fair test of the drug and without understanding of and taking countermeasures to overcome adverse event interferences they experienced with dose escalations [Unpublished data]. These investigators also report no evidence that the sites they engaged were experienced with AD patients and rating scales and other practices and procedures at the core of a successful CT. Each of these limitations might, until shown otherwise, fairly be assigned as posing for the proposed CT a $p=0.50$ risk that the outcome will be determined by errors due to the condition of investigation rather than drug. Just from these probabilities and a chance risk of error statistical analysis model of $p=0.05$, an overall risk of failure of $p=0.76$ argues against continuing with the CT as planned.

Table 5 and Fig. (4) show excess variance and wide inter-site differences in placebo group ratings able to account for the failure of phenserine in the Axonyx CT (AX-CL-06). Subsequently, two small CTs, conducted under tightly controlled conditions, showed evidence supporting efficacy in AD for phenserine as an anticholinesterase [17,42]. Again a checklist can identify issues for consideration by investigators but does not protect investigators from taking irrational bets or choices.

CONCLUSIONS

What are the potential roles for preemptive error management in AD drug developments?

This overview is consistent with the proposition that AD drug developments may currently suffer from self-inflicted wounds. Decisions taken by developers of tarenflurbil, metrifonate, and phenserine, based on the information available to the developers at the time of their

decision, seem irrational with regard to an aim of seeking successful development of a compound as an AD drug. From the data available to us we can not determine whether investigators are unaware of the risks from the choices taken, too inexperienced in drug developments to proceed with the needed caution, forced by business superiors in their firm into decisions they would not otherwise have chosen, driven by circumstances such as the pharmaceutical industry's high value on rushing to market or problems financing developments, or otherwise motivated into what appears, in retrospect, as irrational choices [24]. In the examples we provide and in similar drug developments checklists can be used as a preemptive reminder in planning phases of a development to consider later possible risks and, retrospectively, for evaluating the soundness of progress towards prevention at milestones. Checklist items can raise for informatively prepared and reasonably receptive minds historically identified warnings of pitfalls or risks if preemptive actions are not taken against specific error sources. A checklist cannot, as Zeus warns for human nature, overcome an investigator's "reckless ways...(able) to compound their pains beyond their proper share" [25]. Although we have no evidence other than that presented herein to support our views, we are concerned that, if not scientifically disciplined, neuropsychiatric drug development decisions, when based on opinions or priorities that are other than scientific, risk continued high rates of failures. Our analyses of three failed developments suggest that their failures had roots not in fates, drug properties, or unforeseeable chance events but, in how the drug developments were managed.

One explanation for investigators taking apparently irrational choices that bring on the misery of failed AD drug developments and CTs is a fundamental lack of appreciation for human fallibilities and the magnification of risks from these limitations under conditions of complexity and stress. Drug developers may not adequately appreciate how scientific practices can function as countermeasures to these human limitations. As we argue above, in drug developments, as in the delivery of health care [26], operating nuclear plants, or flying an aircraft, human mistakes are inevitable. In drug developments, unless controlled, the errors that result from these mistakes will almost inescapably be hidden, cumulative, delayed in their effects, undetectable, and often consequential for success or failure. Checklists provide a very limited support to users because, unlike truly user-friendly methods, a checklist item can only remind the user of historical sources for concerns and the need to assess consequences particular to the current situation. Every drug development involves drug-organism interactions and each drug-organism interaction generates its own uniqueness that may, or may not, be consequential for benefits and risks to humans. Vigilance against biological surprises is a necessity when physicians use drugs with research subjects or patients. Checklists are tools to support vigilance, not solutions to problems of errors in drug developments.

Checklists increase opportunities for users to be vigilant and to act preemptively against one of clinical pharmacology's main adversaries in drug development—errors able to invalidate investigators' efforts. By drawing attention to known sources for failures and to situation specific risks, checklists can initiate a process that moves investigators to undertake corrective actions. For example, Kobak *et al.* [27] and Demitrack *et al.* [28] each found that some rater candidates could not achieve sufficient accuracy in ratings to contribute to a CT values reflective of the clinical status of a patient. If investigators do not take preemptive action to exclude similarly unqualified raters from their CTs, then these raters will contribute inaccurate data undermining estimates of efficacy or preventing demonstrations of efficacy [29].

Checklists also promise to expand applications of scientific standards throughout the whole of a drug development from preclinical to Phase III CTs. By directing investigators to introduce, at early stages of planning, interventions against later errors the body of past

experiences associated with checklists structures CTs in ways required by science. Science operates as a body of practices some of which exclude errors from data. Investigators must exclude errors to have confidence that the data obtained in experiments confirms or disconfirms predictions from scientific theory and knowledge rather than conditions peculiar to the experiment itself. For drug developments, checklists aid investigators to expand and organize their efforts such that the development more closely complies with these scientific standards for data and for sequences of reasoning, priorities, and procedures at the core of scientific investigations in other fields.

Acknowledgments

Funding: This work was supported in part by the Intramural Research Program of the National Institute on Aging, NIH. The authors are grateful to Jason Eaton (Senior Visual Media Specialist, Intramural Research Program of the National Institute on Drug Abuse) for creating Figs. (1 and 3).

REFERENCES

1. Aisen PS, Gauthier S, Vellas B, Briand R, Saumier D, Laurin J, et al. Alzhemed: a potential treatment for Alzheimer's disease. *Curr Alz Res* 2007;4:473–478.
2. Alzforum. FDA deems U.S. Alzhemed trial results inconclusive. Viewed April 15, 2010: <http://www.alzforum.org/new/detail.asp?id=1647>
3. Gauthier S, Aisen PS, Ferris SH, Saumier D, Duong A, Haine D, et al. Effect of tramiprosate in patients with mild-to-moderate Alzheimer's disease: exploratory analyses of the MRI sub-group of the Alphase study. *J Nutr Health Aging* 2009;13:550–557. [PubMed: 19536424]
4. Schneider LS, Lahiri DK. The perils of Alzheimer's drug development. *Curr Alzheimer Res* 2009;6:77–78. [PubMed: 19199878]
5. Becker RE, Unni LK, Greig NH. Resurrecting clinical pharmacology as a context for Alzheimer disease drug development. *Curr Alzheimer Res* 2009;6:79–81. [PubMed: 19199879]
6. Wilcock GK, Black SE, Hendrix SB, Zavitz KH, Swabb EA, Laughlin MA. Tarenflurbil Phase II Study investigators. Efficacy and safety of tarenflurbil in mild to moderate Alzheimer's disease: a randomised phase II trial. *Lancet Neurol* 2008;7:483–493. [PubMed: 18450517]
7. 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. [PubMed: 20009055]
8. Doody RS. Dimebon as a potential therapy for Alzheimer's disease. *CNS Spectr* 2009;8(Suppl 7): 14–6. discussion 16–8. [PubMed: 19890242]
9. Miller G. Pharmacology. The puzzling rise and fall of a dark-horse Alzheimer's drug. *Science* 2010;327:1309. [PubMed: 20223954]
10. 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 Alzheimers Dis* 2008;15:303–325. [PubMed: 18953116]
11. 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.
12. Becker RE, Greig NH. Lost in translation: neuropsychiatric drug developments. *Science Translational Medicine*. 2010 In press.
13. Reason, J. *Human Error*. Cambridge University Press; Cambridge: 1990.
14. Becker RE, Greig NH. Neuropsychiatric clinical trials: should they accommodate to 'real-world' practices or set standards for clinical practices? *J Clin Psychopharm* 2009;29:56–64.

15. 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;10:124–131. [PubMed: 8876775]
16. 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;12:54–57. [PubMed: 9539412]
17. 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;63:621–631. [PubMed: 18300284]
18. Butcher J. The hunt for drugs to modify Alzheimer's disease. *Lancet Neurology* 2007;6:1038–1039. [PubMed: 18031701]
19. Becker, RE.; Greig, NH. Methodological Error Risks to Neuropsychiatric Clinical Trials. Viewed April 15, 2010. Available at: www.aristeatranslationalmedicine.com
20. Nordgren, I.; Holmstedt, B. Metrifonate: a review. In: Giacobini, E.; Becker, RE., editors. *Current research in alzheimer therapy*. Taylor & Francis; New York: 1988. p. 281-288.
21. World Health Organization. *Dichlorvos*. World Health Organization; Geneva: 1989.
22. Holmstedt, BR.; Nordgren, I. Toxicological limitations to cholinomimetic therapy. In: Becker, R.; Giacobini, E., editors. *Cholinergic Basis for Alzheimer Therapy*. Birkhauser; Boston: 1991. p. 155-161.
23. Becker, RE.; Moriearty, P.; Unni, L. The second generation of cholinesterase inhibitors: clinical and pharmacological effects. In: Becker, R.; Giacobini, E., editors. *Cholinergic Basis for Alzheimer Therapy*. Birkhauser; Boston: 1991. p. 263-296.
24. Friedhoff, LT. *New Drugs*. Pharmaceutical Special Projects Group Publishing; New York: 2009.
25. Homer. *The Odyssey*. Fagles, R., editor. Penguin; New York: 1996.
26. Committee on Quality of Health Care in America. *To Err Is Human: Building a Safer Health System*. National Academy Press; Washington: 2000.
27. Kobak KA, Engelhardt N, Lipsitz JD. Enriched rater training using Internet-based technologies: a comparison to traditional rater training. *J Psychiatr Res* 2006;3:192–199. [PubMed: 16197959]
28. Demitrack MA, Faries D, Herrera JM, DeBrotta D, Potter WZ. The problem of measurement error in multisite clinical trials. *Psychopharmacol Bull* 1998;34:19–24. [PubMed: 9564194]
29. Cogger KO. Rating rater improvement: a method for estimating increased effect size and reduction of clinical trial costs. *J Clin Psychopharmacol* 2007;27:418–20. [PubMed: 17632240]
30. Becker R, Greig NH. Alzheimer's disease drug development in 2008 and beyond: problems and opportunities. *Curr Alzheimer Res* 2008;5:347–357.
31. Pangalos MN, Schechter LE, Hurko O. Drug development for CNS disorders: strategies for balancing risk and reducing attrition. *Nat Rev Drug Discov* 2007;6:521–532. [PubMed: 17599084]
32. Mangoni AA, Jackson SHD. Age-related changes in pharmacokinetics and pharmacodynamics: basic principles and practical applications. *Br J Clin Pharmacol* 2004;57:6–14. [PubMed: 14678335]
33. Becker RE, Markwell S. Problems arising from generalizing of treatment efficacy from clinical trials in Alzheimer's disease. *Clin Drug Invest* 2000;19:33–41.
34. Becker RE. Lessons from Darwin: 21st Century designs for clinical trials. *Curr Alzheimer Res* 2007;4:458–467. [PubMed: 17908050]
35. 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;26:71–74. [PubMed: 16415710]
36. Lachin JM. The role of measurement reliability in clinical trials. *Clin Trials* 2004;1:553–566. [PubMed: 16279296]
37. López-Arrieta J, Schneider L. Metrifonate for Alzheimer's disease. *Cochrane Database of Systematic Reviews*. 2006;(Issue 2) Art. No.: CD003155. DOI: 10.1002/14651858.CD003155.pub3.
38. Hendrix SB, Wilcock GK. What we have learned from the Myriad trials. *J Nutr Health Aging* 2009;13:362–364. [PubMed: 19300881]

39. International Programme on Chemical Safety. Trichlorofon. World Health Organization; Geneva: 1992.
40. International Programme on Chemical Safety. Dichlorvos. World Health Organization; Geneva: 1989.
41. Axonyx Inc.. New York, NY: Press Release, February 7, 2005:
http://findarticles.com/p/articles/mi_m0EIN/is_2005_Feb_7/ai_n9494083/
<http://findarticles.com/p/articles/mi_m0EIN/is_2005_Feb_7/ai_n9494083/> (last viewed August 10, 2010)
42. Winblad B, Giacobini E, Frölich L, Bruinsma G, Becker RE, Greig NH. Phenserine efficacy in Alzheimer's disease. *J Alzheimer's Dis.* 2010 in press.

Fig. (1). Swiss Cheese model of error [13]. As illustrated, during processes of planning and executing complex processes, investigators and sponsors allow lapses in defenses against mistakes (Active Failures or Unsafe Acts) opening the processes or outputs to errors. These lapses may be oversights or results from effective interventions at multiple levels being overcome by random shifts and realignments that allow, as the figure illustrates, unexpected errors and consequences.

Fig. (2).

Relation between tarenflurbil action in mild to moderate Alzheimer's disease and maximum plasma concentrations of tarenflurbil ($C_{\max}$) in a randomized phase II CT trial [reprinted from [6], Figure 3C]. The straight line of the ascending improvements in the outcome measure at the highest dose tested, expressed as the $C_{\max}$, allow the possibility that more efficacy may occur at greater doses than those within the tested range that were selected for the Phase II CT. All of these tested doses were described as safe and well tolerated, suggesting that tolerability may not have been an issue in not selecting a higher dose.

Fig. (3).

Efficacy and Toxicity with Metrifonate (mg/kg/wk): Weekly Versus Daily Dosing Effects on Percent AChE Inhibition (Assessed in Red Blood Cells (RBC)). Legend: These previously mainly unpublished data are from the author's (REB) laboratory and from [5,15,16]. These data are based on work completed by the author and the monographs [39,40]. The distinction exemplified favors the weekly dosing both because the estimated replacement half-lives of acetylcholinesterase of 4 and 14 days [5] and recovery of other enzymes irreversibly inhibited by the organophosphate pro-drug metrifonate activation [5] are thought to allow recovery by new protein synthesis or activation which does not occur adequately with daily dosing.

Fig. (4).

Variance Sources and Consequences. Inter-site variance in the placebo data derived from the initial phase III CT of phenserine in mild to moderate AD (Axonyx AX-CL-06) (Data from Axonyx master file by permission of QR Pharma). The further variance around the mean values for each site (intra-site variance) is not shown. In light of the expected moderate improvement in ADAS-cog found for anticholinesterases in AD, a large variance in placebo data may over shadow and mask any drug related efficacy signal.

Failures to Validate Methods in Recent AD Drug Developments.(Representative Issues Selected from among a Total of 56 Items Evaluated in [10]. As Stated in the Text, 40 Randomly Selected AD Drug Development Documents (10 in each Category) were Reviewed and the Percentage of Documents Addressing each of the 56 Error Risks Calculated. The Following are Presented as Illustrations of Typical Response Rates to these Error Risks [10])

Table 1

Issue	Literature Reference	Percent Compliance in Study Reports			
		Approved in AD	Failed in AD	MCI	In Development
5. Brain Drug Target Concentration Evidence	Importance of animal & human studies of dosing for concentrations of drug at brain targets [30, 31]	0%	11%	10%	10%
8. Elderly PK/PD effects considered?	Changes with aging potentially affect drug concentrations at molecular targets [32]	0%	6%	10%	10%
29. Rater Qualifications	Kobak [27]: among 29 raters 72% first learned rating scale for the investigation; only 38% had observed using the scale with patients.	0%	0%	10%	0%
32. Ratings' Precision Verification	Imprecision increases the variance in data, undermines statistical significance, and reduces effect size" [33-36].	0%	0%	0%	0%

Table 2

Examples of Promising Drugs Possibly Lost to Human Errors

Drug	Outcome of Development	Errors Able to Account for Failure	Evidence of Drug Efficacy	Comments
Tarenflurbil	Abandoned after failed Phase III CT [7].	Drug concentrations at target not assured [38].	Abandoned. No further work with drug [38].	Errors are cumulative [12]. Hypotheses are lost [12].
Metrifonate	Efficacy in large clinical trials lost to cases of respiratory paralysis [37].	Dosing at levels contraindicated in preclinical work (Fig. 3).	Two positive small clinical trials and efficacy in large trials [15,16].	Non-scientific interests drive development choices [24].
Phenserine	Failure to demonstrate significance in phase III [41].	Excessive variance undermined power [34].	Two positive small clinical trials [17, Unpublished data].	Disease modifying potential not assessed in development [4].

Table 3

Examples of Common Errors: Why We Need Prevention and Control

Error	Effect on Clinical Trial	Where Problem Occurs	Is Problem Detectable?	Is Problem Preventable?	Where Problem is Preventable
Inaccuracy [35]	Numerator distortion undermining power	In CT	No	Yes	Pre-CT with criteria for accuracy
Dose range [12]	Effective drug dose not tested	In CT	No	Yes	Pre-CT with evidence of effects on mechanism in humans
Lax monitoring: entry criteria not met by subjects [27]	Trial report misrepresents subject sample	In CT	No	Yes	Pre-CT with training and In CT with off site evaluators

Table 4

Examples of Pre-Phase III CT Checklist Items

Issue	Resulting Error	Risk Assessment	Acceptable Outcome?
Is Accuracy evidenced for outcome measure(s)? Yes or No?	Inaccuracy of endpoint measures resulting in high variance.	What is the chance that inaccuracy will undermine the validity of the study data set? $P = 0 \dots$	Is this an acceptable risk? Yes or No?
Is Dosing optimized? Yes or No?	Lack of efficacy if doses are too low or toxicity if too high.	What is the chance that dosing is not optimized for the drug and its efficacy can not be tested? $P = 0 \dots$	Is this an acceptable risk? Yes or No?
Will sites comply with protocol conditions? Yes or No?	Enrollment of inappropriate patients not fitting protocol criteria.	What is the chance that sites will not comply with protocol conditions? $p = 0 \dots$	Is this an acceptable risk? Yes or No?
Will investigators comply with prior agreed analysis of primary and secondary endpoints? Yes or No?	Possible introduction of bias with modifications of protocol after study initiated.	What is the chance that the protocol changes will bias the study analyses or interpretations? $P = 0 \dots$	Is this an acceptable risk? Yes or No?

Table 5

Variance Sources and Consequences: Metrifonate vs. Phenserine CTs

Variance Components	Metrifonate * -2 3mo./6mo.	Phenserine * -2 3mo./6mo.
Individual Severity- Fixed	Yes MMSE 10-26	Yes MMSE 8-24
Inter-site- Variable	Not Applicable	Yes Est. SD = 4.5 (129%-MMSE)
Intra-site- Variable	Not Applicable	Yes Est. SD =100-170%
4. Temporal- Fixed	No	No
5. Intervention- Fixed	Var. Increased.	Var. Increased.
6. Random Measurement Error Variance-Variable	+/-2.5 for Means of 3 Observations	+/-4.5 (180%) for Single Observations
Sample Size	50/50	75/215
Drug-Placebo Differ- p = Variable	p = <0.01/<0.03	p = n.s./n.s.
CT Costs- Variable	\$	\$\$\$\$\$-Cases Cost

* The 3 and 6 month CTs for metrifonate and the other ACHEI phenserine were matched for mean differences between actively and placebo treated arms.