



Published in final edited form as:

*J Alzheimers Dis.* 2008 October ; 15(2): 303–325.

## Why Do So Many Drugs for Alzheimer's Disease Fail in Development? Time for New Methods and New Practices?

Robert E. Becker, M.D., C.M.<sup>1,\*</sup>, Nigel H. Greig, Ph.D.<sup>1</sup>, and Ezio Giacobini, M.D., Ph.D.<sup>2</sup>

<sup>1</sup> Drug Design & Development Section, Laboratory of Neurosciences, Intramural Research Program, National Institute on Aging, National Institutes of Health, Baltimore, MD 21224, USA <sup>2</sup> Department of Rehabilitation and Geriatrics, University of Geneva Medical School, Geneva, Switzerland

### Abstract

Alzheimer's disease (AD) drug developments and clinical trials (CT) remain vulnerable to problems that undermine research validity. Investigations of CT methods reveal how numerous factors decrease active drug-placebo group differences and increase variance, thereby reducing power to reach statistical significance for outcome measure differences in AD CTs. Such factors include, amongst many, inaccuracy, imprecision, bias, failures to follow or lack of operational protocols for applying CT methods, inter-site variance, lack of homogeneous sampling using disorder criteria. After a review of the literature and survey of a sample of AD and Mild Cognitive Impairment (MCI) CTs, the authors question whether problems of human error preclude AD researchers from continuing their dependence on rated outcome measures for CTs. The authors propose that the realities of AD, especially a probable irreversible progression of neuropathology prior to onset of clinical symptoms or signs capable of differentiating persons at risk for AD from normal aged, require AD investigators and clinicians to privilege biomarkers and encourage their development as surrogate targets for preventive AD treatment developments, testing, and use in clinical practice.

### Introduction

According to recent research into clinical trial (CT) methodologies, Alzheimer's disease (AD) drug developments and CTs remain vulnerable to lurking problems capable of compromising future research. To determine whether or not research into limitations in CT methodologies are taken into consideration in AD drug developments at present, for the current article we reviewed the literature that we could find with potential relevance for AD CT and drug development problems, developed a list of concerns with potential problems from that review, and then, using this list of potential concerns, we surveyed a representative sample of past and current AD drug development publications for evidence that investigators demonstrated awareness of methodological risks to the validity of their studies.

Four drugs are approved and currently used in AD: donepezil (*Aricept*) 1997, rivastigmine (*Exelon*) 2000, galantamine (*Reminyl*) 2001, and memantine (*Namenda*) 2003. In contrast, we recently identified over 100 compounds tested as potential therapies that were either abandoned in development or failed in CTs [1]. Furthermore, cholinesterase inhibitors (ChEIs), approved for AD, lately failed in CTs to prevent progression of patients with Mild Cognitive Impairment (MCI) to AD. Currently, many new compounds are in clinical testing

\* Correspondence to Robert E. Becker, 2008 Mahre Dr., Park City UT 84098, USA. rebecker2008@comcast.net.

or will shortly enter clinical testing for AD and MCI [2]. Our concerns are that methodological factors, relatively widely reported in the literature as barriers to CT successes, will interfere with investigators providing a fair test for these new AD drug candidates [1,3]. In this paper we consider our own experiences with AD drug development, published reports on CT difficulties especially in neurology, psychiatry, and AD, and we review 40 AD drug developments randomly selected from known AD drug candidates in order to judge whether or not past problems in AD drug development may unnecessarily impede research into the efficacy of drugs either now in CTs or about to enter clinical development.

Since the 1984 publication of NINCDS-ADRDA criteria for diagnosing AD, progress understanding relevant genetics, imaging, biochemistry, and clinical presentations justifies revisions to AD diagnostic criteria [4]. Recently, Pangalos et al. [5] described advances in drug development practices over the last two decades. The potential importance of these technological advances is suggested by the latest Institute of Medicine [6] review of studies used to formulate clinical policies for treatment of post traumatic stress disorder (PTSD) [7]. The IOM Committee on Treatment of Post Traumatic Stress Disorder found 2771 studies of PTSD but evaluated only 90 as sufficiently sound scientifically to include in their review. They found the evidence inadequate in these reviewed studies to support the efficacy of any psychoactive drugs or psychotherapies for PTSD other than exposure therapy. The IOM Committee cited numerous problems with sample size, blinding, independence of investigators, the need to exclude studies that were case studies and case series, high dropout rates capable of introducing bias, influences associated with pharmaceutical funding, publication bias, population selections, and other factors [6, pp 124-125, 136]. After ascribing some problems to earlier studies not using current investigational standards, to overcome the “scientific inadequacy” of available CTs in PTSD, the IOM report recommends that investigators use “the latest and most rigorous methods for designing and executing study protocols”. A Type II error occurs with the mistaken dismissal of a research hypothesis of drug efficacy because study conditions failed to provide a fair opportunity for the drug to express its true effects [3]. In the current paper, we ask whether, by currently taking advantage of research into methodological limitations that can lead to Type II errors in interpreting research studies, AD drug development exemplifies the progress in clinical research expressed in the IOM report.

## Background

### Overview of concerns with AD drug development

Currently, investigators focus AD and MCI investigations on disease modification rather than symptom remission. In AD drug development, the CT, originally designed for short-term evaluations of treatments, has not yet been successfully modified to give investigators confidence that disease modifying drug effects can be successfully tested [8-10]. Long duration trials with non-parallel initiations and terminations of treatment to detect disease modifying drug effects risk increased losses of subjects and other compromises to their integrity [11]. It is widely accepted that designs able to test disease modifying drug effects will require larger numbers of subjects, greater numbers of clinical research sites, and longer double-blind evaluations [12]. Under these conditions, inter-site variance, difficulties monitoring sites for conformity to protocols, increased losses of subjects, placebo responses, and life-events more easily undermine CTs [13,14].

Pangalos et al. [5] point out that the rate of AD drugs coming to market falls below the already low 7% rate for all central nervous system drugs. Consistent with the IOM report on development of treatments for PTSD and with Pangalos et al., we and other authors raised, for clinical phases of neurological, psychiatric, and AD drug developments, serious

questions about the reliabilities of development methodologies and practices [1,15]. As Esterbrook et al. [16] and others encountered as they attempted to evaluate development programs for drugs, many unsuccessful preclinical and clinical studies are not reported in the literature and results will not be released by sponsors [17]. Occasionally, a synopsis that lacks considerable detail may appear as a press release [1]. These long-recognized problems exerts an unknown effect on our understanding of how methodologies and practices may contribute to drug development failures in AD [1,5,18].

### **Potential weaknesses affecting the validity of AD drug developments**

Investigators find problems of inaccuracy, imprecision, and bias interfering with CTs that are dependent on using clinical ratings as outcome measures [14,15,19-25]. Error components, added to clinical ratings by inaccuracies, imprecision, failures to follow or lack of operational protocols for applying methods, and bias, have been shown to decrease active drug-placebo group differences and to increase the variance in data. These effects on data reduce the possibilities for reaching statistical significance for outcome measure differences [3,15]. Similar measurement errors and a lack of specificity during diagnostic evaluations and qualifications of subjects for eligibility for a CT can include subjects incapable of responding to treatment because of misdiagnosis, genetics, or specific pathology [25-26]. Each of these factors reduces the power of the CT to detect drug-placebo differences; requiring additional subjects, additional sites to provide subjects, more investments in training investigators, risks of recruiting clinically inexperienced investigators, increased resources for monitoring sites thus contributing further risks of compromising the integrity of the CT. Studies have shown how inaccuracy and imprecision in outcome ratings cause CTs to fail from compromises to power and from complications that follow from additions of subjects to increase power [3,12,24,25,27-30]. Unreliability, due to limitations in the sensitivity of outcome measures to change and human errors, risks false-negative reports and patients being exposed to inadequately designed and controlled research conditions [1,31,32].

### **Problems with raters**

Engelhardt et al. [15] found that over 50% of clinical evaluators in two multicenter trials of antidepressant drugs functioned poorly or only fairly as evaluators of outcomes. Cogger [27] demonstrated that these poor and fair raters reported no active drug-placebo group differences. Rather, only good and excellent raters, in terms of compliance with protocols for using the Hamilton Depression Scale, contributed to the effect size associated with the antidepressants [27]. Kobak et al. [33] determined that among 29 raters at 12 antidepressant study research sites 72% first learned to administer the Hamilton Depression Scale [34] in preparation for the investigation and that only 38% were observed using the scale with patients. Demitrack et al. [35] found excessive variance among sites in large multisite studies. Kobak et al. [36,37] and Targum [38] showed beneficial effects of more intensive training of investigators at research sites. On the other hand, Demitrack et al. [35] and Kobak et al. [37] found that, for some prospective raters, extended training did not improve the reliability of their performances necessitating their exclusion from participation in CT research. There is no agreement over how to overcome problems of rater unreliability although, as a result of recent studies documenting the problems, investigators are debating means to overcome these limitations to CT validity [39,40].

### **Problems of inter-site variance**

Consistent with these concerns raised by researchers into CT methods, two highly experienced developers of AD drugs, Paul Aisen and Bruno Vellas, recently acknowledged the seeming inevitability in AD studies of “large degrees of variance amongst sites,” the large number of subjects requiring “70-80 sites” dispersed “around the world,” and sites

contributing only “2-5 patients.” [12]. Cummings [41] claims that clinical evaluations using rating scales can be used as biomarkers and qualified as surrogate measures for AD disease severity. We and others present evidence of problems introduced by excessive variance from the human diversity of subjects, inherent imprecision in rated scale values, and shortcomings of evaluators prone to human errors and bias. This evidence suggests that clinical evaluations are at best too imprecise and inaccurate to overcome the need for large numbers of subjects to power CTs and the resulting complications from large numbers of subjects noted by Aisen and Vellas.

### Problems with rating scales

AD rating scales are limited to ordinal data. Ordinal data express a sorting of subjects in relation to others or selves at different times; but, without any specific intervals expressed in the data. Unlike a change from 7 to 8 centimeters where the size of the interval is specified, for a subject to remember 8 rather than 7 words implies no specific quantity that separates the two categories. Rating scales can indicate but not quantify differences. As Winbald et al. [42] worry, “small numerical changes on scales...might...be clinically meaningful in practice” and vice versa. Effect size estimates, because they use qualitative and not quantitative data, do not entirely overcome these concerns. Recently investigators ascribed to time effects on ratings trends towards smaller rates of annual decline in AD patient groups demonstrating the latent instability of qualitative rankings [11]. The movement towards disease modifying interventions requires early diagnosis and evaluations of clinical status when only initial cognitive and behavioral effects from AD are manifest and possibly even earlier. Existing AD ratings scales exhibit ceiling insensitivities when used with pre-AD conditions. These insensitivities lead investigators to turn to more sensitive outcome measures such as the Neuropsychological Test Battery [31]. As we have noted [1], many rating scales and tests are insensitive to changes in early or pre-AD. Using these scales minor rated differences in pre-AD may not reflect much greater and irreversible changes in underlying pathology. This insensitivity to irreversible pathology suggests the importance of quantitative biochemical markers for identifying pre-AD as a disease process. Ordinal values obtained using rating scales are associated with relatively large inherent variances. This variance expresses imprecision in measurements and limitations when raters attempt to quantify changes in a patient's clinical state. Because of these limitations ordinal data from rating scales can not provide reliably sound grounding for extrapolating from rated clinical changes to neuropathological changes, especially in pre-AD states. AD preventive efforts, to be effective, may have to be applied prior to appearance of clinical disturbances creating serious evaluation problems for researchers and management problems for practitioners without reliable methods of clinical assessments for pre-AD.

Becker and Greig [1] and Becker [3] discussed how CT methodologies are made more problematic by the dependence of AD drug developments on rated clinical outcome measures. Estimates of sample sizes from studies of biomarker changes over the course of AD put into dramatic perspective the magnitude of the complications introduced into CTs by dependence on human judgments and rated outcomes [11]. Thal et al. [11] estimated that numbers of subjects needed per group in AD CTs could be reduced from the current range of 270-6377 (Table II) to 16 to 40 based on what is currently known of imaging and biochemical markers.

### Establishing dosing ranges, dose titrations, and effects on dosing from individual pharmacokinetic-dynamic diversity—difficult and crucial problems

Becker and Greig [1] and Pangalos [5] discuss recent CTs where patients were dosed below the dosing ranges shown effective in earlier Phase II dosing studies and confirmed in CTs. They emphasized the importance of establishing dosing parameters in early clinical studies,

having evidence from animal studies that these parameters provide appropriate concentrations of drug at brain targets, and insuring that these dosing protocols control later CT dosing and drug management in clinical practice. Becker et al. [19,20,52] found dosing by weight necessary to approximate an identified optimal dose. Microdosing studies provide one effective way to confirm drug concentrations at brain targets otherwise open only to extrapolation from models developed *in vitro* and in animal studies [1]. Dosing studies must also take into account problems of titration and adherence that may occur in later clinical practice [43-46]. Bellelli et al. [47] found that 59% of community practitioners did not increase ChEIs from their lowest starting doses. A range of studies may be needed to adjust dosing for implementation under real world conditions and for the well known changes with aging that potentially affect drug concentrations at molecular targets. These changes include fat compartment increases, water compartment decreases, serum albumin decreases,  $\alpha$ 1-acid glycoprotein increases, altered hepatic P-450 enzyme system metabolism, decreased renal clearance, and so forth that impact pharmacokinetics, pharmacodynamics, as well as potential drug-drug interactions.

### Drug kinetics and dynamics not adequately considered in drug developments

From what is published one can speculate that sponsors of drug developments in AD do not take advantage of traditional stepwise research preparing for CTs or use to full advantage pharmacokinetic and pharmacodynamic resources. It appears that in a rush to market Myriad, Neurochem, Sanofi, Elan, Bayer, and other pharmaceutical firms use post-hoc analysis, animal and not human dosing testing, biochemical marker changes, business priorities, and Phase III without preliminary evidence for dosing and efficacy at Phase II and earlier stages [1]. The authors found drug kinetics and dynamics critical to effective use of metrifonate and have argued elsewhere that disregard for basic pharmacokinetics and pharmacodynamics led to the failure of metrifonate to gain an NDA [1,48-52].

In the literature we could not find extended discussions of the prevalence of or consequences from research sites falsifying data, lack of adherence of raters and administrators to protocols, investigators at sites lax applications of diagnostic criteria to secure subjects, the possible impacts of advertising for patients for a CT compared to patients who present spontaneously for diagnosis. We encountered each of these problems during participation in multicenter CTs, over the last 40 years confirmed their presence in informal discussions with colleagues and study monitors, and occasionally read of their importance [12]. The literature indirectly offers support for consideration of these issues, for example, Kobak et al. [25] find site raters exaggerating pathology apparently to qualify subjects for research studies, Carroll [53] views pharmaceutical firms ignoring subtypes of depressive diagnoses in CTs to gain unrestricted drug approvals, and Petersen and Morris [54] point out the increased heterogeneity in unselected epidemiological subject samples compared to those who seek help from a care provider for a problem, a difference that may also occur between self-selected patients and those persons responding to advertisements for research subjects.

Kelloff et al. [54] describe in cancer studies, and Pincus and Stein [10] in rheumatoid arthritis studies how inappropriate populations and settings prevent expressions of drug efficacy raising the risk of Type II interpretations of CT outcomes. While, with Raschetti et al. [55], these authors emphasize the importance of criteria to regularize practice across research sites, they do not call specifically for development of protocols to systematize, control, and replicate conditions and practices across centers or for investigators in centers [11]. Williams' [56] protocol for administration of the Hamilton Depression Scale allowed Engelhardt, Kobak, and their colleagues to uncover the inaccuracies introduced by outcome rating practices and in qualifying candidates as subjects for depression CTs [15,25]. They took important steps towards operationalizing the concept of the careful interview Petersen

and Morris [54] depend on for diagnostic validity. As we discuss below, failures to operationalize MCI criteria may, in part, account for a lack of success in those CTs [57].

### **Bias in test administration expressed in placebo controls**

Imprecision in measurement can be expressed as the standard deviation of repeated test-retest administrations of a test under conditions that should not affect a subject's or measuring instrument's performance [21,58]. Becker and Markwell [21] demonstrated that the standard error of measurement for test-retest data is a more sensitive indicator of potentially interfering error than are reliability coefficients. Inaccuracies in measurement require comparisons with evaluations produced by 'gold standard' raters or instruments. A further problem in our experience, but not for Ferris [59], is bias expressed in AD CTs by placebo group improvements on outcome measures over the period of the double-blind [3]. Becker and Greig [1] conclude that the demonstrated magnitudes of imprecision, inaccuracy, and bias found in AD and other CTs precludes long-term dependence of AD investigators on clinician rated outcome measures as endpoints for individual CTs. On the other hand, investigators such as Schneider [60], look to frequent use of improved sensitive neuropsychological tests to overcome difficulties with current CT practices. Becker and Markwell [21] demonstrated that using the mean of ratings, rather than individual ratings, improves precision and power to detect statistically significant differences between treatment arms [3]. Kobak et al. [37] and Targum [38] demonstrated improved accuracy with extended intensified rater training. The effectiveness of each of these methods depends on the manner in which they are used. In studies using neuropsychological tests, biochemical assays, imaging and, even sporadically, in applications of diagnostic criteria and uses of ratings, investigators specify with protocols the conditions for use of the test [11,54,56,61,62]. Generally, extensive protocols to control practices are not provided in CT proposals or reports.

### **Overview of problems in MCI drug development**

Petersen [57] observed how study design vulnerabilities due to heterogeneity of the sample, insensitivity of outcome criteria for progression, subtherapeutic dosing, and long trial length are able to account for the trial's failure. Jelic [109] cautioned that heterogeneous samples at entry to an MCI CT, lack of biomarkers to enrich samples with subjects at risk of deterioration, suboptimal durations of treatment, and ineffective indicators of disease progression must be overcome in future MCI CTs. Harrison et al. [31] developed the Neuropsychological Test Battery to overcome the ADAS-Cog's insensitivity to change in MCI. This test is not practical for use in other than specialized clinics. With the evidence of neuropathological changes probably irreversible prior to appearance of adequate symptomatology or brain volume changes to differentiate pre-AD from normal aging, AD research may have to shift from clinical ratings to practical screening biomarkers to best prevent disabilities from AD. Diagnosis that depends on tissue loss, such as hippocampal atrophy, offers only secondary, not primary prevention and accepts, does not anticipate and prevent, irreversible changes [68-72]. Given the neuropathological change prior to rated cognitive changes and the imprecision of the ADAS-Cog and MMSE, practitioners' clinical judgments, with or without support of testing, are probably not adequately reliable to ground either AD or MCI patient management decisions [14]. A truly most effective preventive intervention for AD will most probably require patient management based on results from biomarkers sensitive to pre-clinical changes predictive of later AD.

### **Setting drug development priorities**

In compiling a list of AD and MCI drug candidates, Becker and Greig [1] found that most AD drugs, with the exception of some traditional remedies such as ginkgo bilboa, were studied with reference to a specific mechanism of action, for example acetylcholinesterase

inhibition, Beta sheet breaking, gamma secretase inhibition, and so forth. Even though AD drug candidates may be active through over 30 different mechanisms, only ChEIs and memantine, an NDMA receptor glutamate antagonist, received US Government Food and Drug Administration (FDA) new drug approval. For example, after adding to the list used for Becker and Greig [1] we easily identified 14 ChEIs that reached clinical testing in AD with 5 approved, 2 under development currently, and the remainder abandoned during development. Becker and Greig [1] discuss the apparent errors in development that appear to have caused the failures of two drugs while others, such as heptylphysostigmine, failed for toxicity of the molecule separate from the mechanism of inhibition. We identified 3 glutamate antagonists with 1 approved and 3 abandoned; 15 muscarinic agonists with none approved and all abandoned, and 12 nootropics, 9 anti-inflammatory drugs, and 4 choline precursors failed in CTs or earlier in development.

It seems generally accepted that ChEIs, as a class, share a mechanism of action potentially beneficial in AD in the absence of toxicity due to the molecule [51] while, with the possible exception of glutamate antagonists, the mechanism(s) of action for each of the other classes, muscarinic agonists, nootropics, anti-inflammatories, precursors of choline in studies to date demonstrate no functional benefits in AD. Given the limited professional and financial resources available to support clinical AD research and the ethical problems of exposing patients to compounds not well justified for further research, we would expect CTs testing compounds, thought to act with previously failed mechanisms, to be provided specific justifications for why this compound should be regarded as a possible exception.

### Subject samples

Raschetti et al. [55] reviewed AChEI CTs in MCI. They found samples being not equivalent across studies impairing cross-study comparisons and potentially contributing to failed outcomes. Given Kobak et al.'s [25] evidence that evaluators rated depressed subjects at baseline more severely than outside observers blind to the situation, it is possible, for example, that Feldman et al. [108] recruited less severely challenged MCI subjects from investigators being under pressure to fill quotas and from problems with entry criteria. Clinical skills affect evaluation abilities [33,63]. Study sites and investigators need training and evaluation in use of entry criteria, including AD diagnostic criteria. Visser et al. [115] found different samples drawn from the same population in their clinic after applying entry criteria used for different MCI studies. Varma et al. [26] demonstrated that the homogeneity of AD research subject samples is inevitably compromised by the inability of NINCDS-ADRDA criteria to exclude frontotemporal dementia cases.

### Biomarker research

Currently magnetic resonance imaging (MRI), positron emission tomography (PET), single positron emission computerized tomography (SPECT), and spinal fluid biochemistry offer promise as potential sources of quantitative biomarkers for AD disease progression. PET, SPECT, 1H MRS, and MR volumetry of the hippocampus can distinguish AD patients from elderly normals [68]. Hippocampal volumes in MCI have been shown to predict progression to AD [69-74]. Longitudinal trends in brain images have been associated with cognitive decline [68]. Nordberg [75] interprets these preliminary successes as showing promise for imaging to support diagnosis and indicate disease severity in AD. Acceleration along this path can be expected from the multi-center coordination of development with the Alzheimer's Disease Neuroimaging Initiative (ADNI) [62]. Currently the methods of choice for imaging AD pathologies have not been identified [76].

Hippocampal atrophy, more prevalent in patients disposed to later AD and in AD patients than in normal aging, correlates with disease severity, but not adequately to show drug

effects on disease progression in periods of less than one year [77-79]. Longitudinal cortical grey matter atrophy in AD witnesses to the need for early diagnosis and treatment [78,80-82]. PET, able to indicate reduced regional cerebral metabolism in AD with fluorodeoxyglucose (FDG) imaging, is useful in AD diagnosis and can assist in excluding frontotemporal dementia [83-87]. Pittsburgh compound B (PiB) and FDDNP allow visualization of amyloid [83,86]. Stable PiB retention after 2 years supports early amyloid accumulations in AD possibly prior to declines in glucose metabolism and cognition [88]. Unfortunately no imaging techniques yet predict for an individual case [89].

In AD amyloid  $\beta$  ( $A\beta$ ) CSF levels decrease, tau fractions increase and BACE1 levels in MCI predict increased risk for developing AD [72,90]. Phosphorylated tau correlates with neurofibrillary tangles (NFT) and with hyperphosphorylated tau concentrations in post mortem AD brains [91]. Unlike  $A\beta$  concentrations that remain stable, t-tau increases from early to advanced AD [92-94]. Hansson et al. [95] proposed the  $A\beta$  42/ $A\beta$  40 ratio as a predictive biomarker for AD [96,97]. Given the rapid progress recently applying imaging and molecular biomarkers to our understanding of AD, it seems reasonable to evaluate whether or not the move to biomarkers can, in the near future, overcome current impediments to most effective uses of CTs in AD drug development. The FDA anticipates that biomarkers will become increasingly important in all phases of drug development [98]. Authorities urge caution using markers to infer clinical outcomes [99-101]. Since the FDA Modernization Act of 1997 allows for fast track drug approval when a surrogate marker indicates a drug as most likely to provide clinical benefit for serious and life-threatening diseases, markers may come to play an important role introducing new AD drugs [67]. As Temple [64] emphasizes, a validated surrogate marker must demonstrate long-term safety in addition to confirmations of predicted efficacy.

As we already noted, hippocampal volumes in MCI have been shown to predict progression to AD. BACE1 levels in MCI predict increase risk for developing AD [90]. Brys et al. [110] found P-tau was the strongest among the CSF biomarkers they investigated for predicting decline from MCI to AD. Combined measurements may show promise for quantifying drug effects on pathological processes in AD and pre-AD [111-113]. Given the failures of drugs in MCI, MCI both should not be treated as a clinical entity and can not at this time be treated creating a barrier to early interventions in AD [114]. On the other hand, research suggests that investigators will develop biomarkers with greater predictive capacities for quantifying progression, targets for treatment that hopefully, as surrogates, will overcome current unreliabilities and the complications for CTs and patient care.

### **Proposed modifications to methodologies: developing and using new CT technologies**

Based on the range of difficulties designing, executing, analyzing, and interpreting AD CTs, Becker and Greig [1] proposed that investigators may, in choosing how they develop AD drugs, fail their drugs by not providing fair tests for efficacy and adequate safeguards against compromises to safety [3,57]. They concluded that limitations inherent in AD drug candidates do not necessarily account for all failures of CTs to confirm efficacy [1,3]. Because of dependence on rating scales, the risk of Type II error may be more prominent in AD drug development and CTs and in other areas of neurology and psychiatry where investigational outcomes are vulnerable to human errors. In accord with Pangalos et al. [5], we suspect methodological factors account significantly for the low success rate of AD drug developments.

In response to the difficulties they and others uncovered, Pangalos et al [5] and Becker and Greig [1] proposed modifications in drug development and CT practices designed to reduce methods and practices interfering with demonstrations of efficacy or other outcome aims. Evidence indicates that some methodologies and practices may be flawed in AD drug



development. Currently the evidence is only circumstantial and more research is required into specific problems: random and systematic errors; bias; the limitations on sensitivity and specificity of rating scales; and limitations on human abilities to use rating instruments with accuracy and without bias; and other factors. It seems unlikely that research can overcome the problems that arise from exclusive dependence in CTs on rated outcomes. Alzheimer Disease Neuroimaging Initiative investigators are currently researching imaging and biochemical biomarkers, their use as potential surrogate outcomes, and protocols for effective applications in research and patient care [11,61,62]. Similar concerns for the reliability of clinical assessments arise in clinical practice where, without rigorous training in evaluation, often with limited skills and experience with AD, and commonly without using established rating instruments, practitioners reach diagnostic and management decisions. Accuracy in diagnosis and management assessments varies with the experience of the rater [33,63]. Problems with diagnostic and rating scale evaluations will not be automatically corrected by advances in rating and testing methods in CTs. For example, the Neuropsychological Test Battery, required to overcome sensitivity barriers in MCI research, where disease modifying interventions will become available for practitioners if research is successful, is impractical in the clinic because of time and skills needed to perform the tests [31].

As a resolution to the difficulties they found inherent in AD CT methods and practices, Becker and Greig [1] proposed intensive development of quantitative biomarkers able to provide molecular outcome targets for AD CTs. They reasoned that, although the FDA appropriately requires demonstrated clinical efficacy in two areas for use of a CT in support of a new drug application (NDA), limiting outcomes in this way, the risks of Type II errors from AD clinician rated outcomes are too great. Using quantitative and pathology-specific biomarkers, such as markers for amyloid, tau, cell death, and so forth, Becker and Greig anticipated the possibility of first confirming molecular effects from an AD drug and then exploring the molecular effects or generalizing from other evidence of clinical benefits associated with similar molecular changes to establish the relationship to clinical benefits. Separate trials, focused on clinical benefits associated with specific modifications of pathology, could then be used to confirm that the CTs with surrogate markers comply with FDA NDA required clinical benefits. Under this proposal a drug may become a useful pharmacological probe or indicator of a class action even though it does not itself offer clinical benefits.

Becker and Greig [1] proposed for consideration three new procedures meant to overcome or reduce risks of Type II errors in interpretation of AD CT results, to reduce interference of real world conditions with determinations of efficacy of a candidate AD drug, and to use a CT as a model defining the conditions of clinical use of the drug required to maximize the opportunity for the drug to be effective with a practitioner's patient. The proposed new procedures involve modifications to currently preferred intention-to-treat (ITT) based analyses of CT data meant to increase the effectiveness of CTs testing for drug efficacy, wider uses of protocols to govern CTs and clinical practices to assure scientific accuracy, precision, control, systematization and reliability, and rapid development of biomarkers as quantitative surrogate markers [9,64-67]. Raschetti et al. [33] and the Alzheimer's Disease Neuroimaging Initiative investigators, Foster [61], Thal [11], and Weiner [62], find protocols necessary to control problems of human and machine errors intrusions into AD research.

## Objectives

Since 1997 the authors have studied how better scientifically to systematize and control AD CTs and patient care. Consistent with Kralawish's recent *NEJM* editorial [102], the authors

emphasized the role of CTs informing patient care and investigated methods to strengthen that information transfer function of AD CTs [1,107]. Given the methodological and practical issues that potentially interfere with AD drug developments providing reliable tests of drug efficacy and that open risks of Type II errors in interpretation of research outcomes, for this paper we sought to first, confirm the presence of methodological and practical problems in a selection of studies taken from AD drug developments and, second, to bring together our own and other's recommendations for changes able to overcome these problems and the interference of problems with AD CTs applying to individual patients in clinical practice and informing health care policy [103-106]. This paper aims to estimate from a small random sample of the literature, the presence or absence of possibly critically important methodological deficiencies able to interfere with AD drug development successes and to justify corrections to methods and practices able to overcome this interference.

## Methods

For this study we organized our assessments of drug development decision making, methods, and practices under six broad activities in drug development: 1) mechanisms of action, pharmacological activities, design of studies and publication; 2) dosing; 3) research subject samples; 4) outcome measures; 5) research sites and investigators; and 6) protocol controls over methods and practices and clinical information transfer.

These topics were chosen from issues raised in the Background and used to guide our analysis of selected AD drug developments. One major problem, encountered by anyone wishing to evaluate why some AD drugs succeed and others fail, is that most negative outcomes of drug developments are not published and sponsors refuse to release information about the development [1,5,16-18,55]. Consequently, in earlier publications [1,3,14] we retreated to analyses of selected reports, personal experiences with drug developments, and other available information that seemed relevant. In this study, using random sampling, we planned to estimate more precisely if and how extensively not publishing investigational outcomes interferes with advancing AD research and patient care.

To update our list of AD drug developments used in Becker and Greig [1] we searched Medline, the Cochrane Collaboration files, the Cochrane Collaboration Trial Register, the Food and Drug Administration controlled clinical trials registry, Current Controlled Trials, clinicalstudyresults.org, clinical trial result postings by sponsoring drug companies on their websites, references in publications, review articles, and investment banking firms' and other published lists [116] of AD drugs considered for investigation or commercial development. We then used a random number table to select drugs and specific studies of these drugs to analyze. After a drug was randomly chosen for inclusion as one of the 10 drugs to be addressed in a category, we randomly chose, whenever possible, among pivotal studies or later reported studies of the drug. When a pivotal study report or study publication from the last reported or, for drugs in development, current phase of development, was not available, we first chose a registry report of results from the appropriate CT, second the registration as a CT, and then third the most recent published investigation of the compound - randomly chosen from a group when more than one publication was available. As a result we realized we would underestimate the rate of failure to publish negative trials in the interests of sampling a range of studies in each of four categories. We developed four separate randomly selected lists, each with 10 papers providing information on drugs: 1) AD drugs approved and used in clinical practice; 2) AD drugs reported failed in drug development or presumed abandoned because of no published references to outcomes from earlier reported research intentions; 3) MCI drugs failed in drug development, and 4) drugs currently under development for AD or MCI.

From our survey of the literature and extrapolations to clinical research from relevant good laboratory experimental practices we then developed a list of 56 specific queries covering the six issues where our survey of the Background literature opened possible methodological impediments to AD drug development. Each of the 56 queries was worded to investigate a specific issue raised in our survey of background literature (See Tables II:1-6). One of the authors rated each selected study for whether each of the 56 topics is reported in the publication selected for inclusion in this study and for how the report in the publication addresses the subject matter of the topic. The other authors reviewed the ratings.

Studies were judged using each query based on whether or not the publication discussed, or at least acknowledged, the specific subject of the query. This produced for each category the percentage of publications in which the subject area was addressed or at least noted. Where data could be compiled, such as numbers of subjects (Table II: 5. Query 39) or numbers of sites (Table II: 5. Query 41), this data was recorded and summarized for each of the four categories of study outcomes.

## Results

### Publication of drug development outcomes

We randomly chose, when available, 10 publications from each of our four categories of drug outcomes. We had to survey the literature for 18 compounds to obtain 10 published reports or acknowledgements of outcomes for unapproved compounds previously reported as being developed for AD. The drugs and references reviewed are listed in Table I. After the random selection was completed we noted Aisen et al. [148] was mistakenly assigned under MCI trials. We retained the trial because the study addresses possible differences between NSAIDs reducing risk of subsequent AD in MCI and stabilizing AD which we saw as relevant to study subsequently of anti-inflammatory effects on MCI syndrome mechanisms of progression.

Our selection gives us an estimate of 56% of compounds, claimed to be under development for AD, without any information available about outcome. We found publications of results at the level of the last reported stage of development for only four compounds of the 18 failed AD compounds, Vitamin E [117], propentophylline [118], ginkgo biloba [119], and alfoscerate [120]. None of the reports identified itself as a pivotal study for the stage (Table I). This provides us an estimate of 22% of compounds in our AD development sample as having some data available from the last reported stage of development and possibly no data about specific studies that led to abandonment of the drug (Table II: 1. Query 11).

For AD approved drugs we found 9 or 90% of randomly selected CTs, published. For MCI 70% of our randomly selected trials were published and for drugs currently under development, except for registrations of CTs, no publications about the current stage, as could be expected since the latest stage of development remains in progress. Assuming that unpublished failed AD drug investigations are negative and, when not referenced in the literature that they underlie abandonment of these drugs, we have only 7 negative pivotal publications in MCI, an estimated 24% publication rate for negative outcomes in AD drug development—1 negative study for an AD approved drug [121] 18 failed AD drugs and 10 failed MCI applications.

### Attention of investigators to potential areas of methodological interferences with development

In the total of 40 drugs with some literature on outcomes of development we found 0% descriptions of methods and procedures in each of the four categories of our sample for 14 out of our 56 queries of trial reports (Tables II:1-6), a maximum of 20% to >/=0% reporting

in 22 of 56 queries, and a maximum of 50% but greater than the range 20% to  $\geq 0\%$  for 2 queries. This estimates that less than a third of the areas of methodological concern we raise ever receive attention in more than 50% of AD investigations. In fact, higher compliance is often reached in only one out of the four outcome categories, suggesting much lower overall levels of investigator concern with the potentially problematic methodological issues that we and others have identified.

### **Establishing and adhering to a well-evidenced effective dosing range**

For the 10 queries we developed to reflect commitments to sound dosing development and practices (Table II: 2), only stating the basis for dosing increases in a CT (Table II: 2. Query 17) and reports of the percentage of subjects reaching target doses (Table II: 2. Query 18) ever exceeded 50% compliance. Six queries (Table II: 2. Queries 12-16, 19) were addressed in 20% or less of the reports we reviewed.

### **Characteristics of research subjects**

Measures to insure sample homogeneity within a study and comparability of study samples with samples drawn for other studies were reported in less than 50% of failed AD and AD in development descriptions (Table II: 3). Although we found criteria listed for each of the 10 MCI studies, the variations among criteria sets, lack of operational definitions [55], and the non-equivalent samples drawn by Visser et al. [115] using the different criteria sets demonstrates how our requirement of attention or description does not tap into the full vulnerabilities of CTs to methodological deficiencies. Consideration of effects on outcomes from other diagnoses and medications were described in 20% or less of the papers and imaging, although required in 20-30% of studies, specifically addressed confounding of the sample by fronto-temporal dementia (Table II: 3. Query 26) and confirmation of AD (Table II: 3. Query 27) in 0% of reports.

### **Attention to rater reliability and effects from unreliability on sample size, numbers of sites, and so forth**

Except for reporting the outcome measures used (Table II: 4. Query 28), none of the other 8 queries (Table II: 4. Queries 29-36) addressing accuracy, precision, or bias as sources of error exceeded 20% consideration in study reports.

One hundred % of papers reported the number of subjects in the study (Table II: 4. Query 39). Information about sample sizes is compromised by the unpublished failed AD studies. Twenty to 70% of papers identified subjects lost during the study (Table II: 4. Query 40) and, leaving out unreported failed AD studies, only 60 to 90% of publications reported the number of involved sites (Table II: 4. Query 41). No reports provided information on the number of sites providing less than 6 subjects to the study (Table II: 4. Query 42). Criteria for qualification and disqualification of sites (Table II: 4. Queries 43-46) were reported by 0% and monitoring activities (Table II: 4. Query 47) addressed in only one paper.

### **Research sites and investigators**

Rater qualifications and site monitoring are reported in Table II: 5.

### **Protocols to control methods, practices, and conditions**

Zero to 10 % of papers described protocols to control the manner in which tests, rating scales, criteria for diagnosis, recruitment, biomarkers or other features of the investigation are applied, used, and interpreted (Table II: 6. Queries 52-53). Studies reporting protocols almost exclusively addressed methods for neuropsychological testing or biomarker assays.

## Information transfer to practitioners

Zero % of studies described methods to transfer methodologies, practices, protocols, and procedures (Table II: 6. Queries 54-56) from the CT to the practitioner.

## Discussion

Science presumes free sharing of information as essential to orderly progress in its knowledge, an ideal more easily realized with current access to information technologies [Young 122]. Our sample of information on outcomes of 48 compounds identified as under consideration as potential AD therapeutics supports earlier claims that drug developments with negative outcomes go unreported. This lack of information about failed drug developments blocks assessments such as we have undertaken to evaluate how methodologies might contribute to drug failures and similar inquires into the pharmacology and therapeutics of AD drugs. As we have proposed elsewhere [1] and as Kralawish [102], Chalmers [18], and others have protested for decades, fair evaluations of the literature on AD drug therapeutics and clinical pharmacological methods are blocked by the biases introduced by the selective reporting of results from scientific pharmacological investigations in AD. Clinical investigators often must work in the dark as commercial sponsors withhold basic pharmacological information about a drug under investigation as proprietary [12]. We suggest that our data add to the body of reporting that calls for core revisions in AD researchers' and practitioners' access to results from investigations. The establishment of clinical trial registries meant to improve reporting from negative outcome investigations is not succeeding according to our data.

Science presumes conditions of investigation that are controlled and described such that other scientists will be able to replicate the original investigation. In our survey of the literature on potential methodological compromises in AD CTs, we find, in the failures to report methods to control possible interference from sampling of subjects, protocols for applications of criteria for admission to a study, consideration of unreliability due to problems with precision, accuracy, sensitivity, selectivity, investigator bias, and other factors. Neuropsychological, imaging, and biochemical studies alone in our sample provided the detailed descriptions of methodologies and conditions for applications of methodologies that allow readers to evaluate the control of conditions in an investigation and provide information needed to replicate the investigation. In our sample of 48 candidate AD compounds we found very low rates of reported attention to the majority of factors highlighted for us from our survey of the literature on methodological risks to validity and our own research and experiences. Within discussions of studies that failed or proposals for investigations we found few, if any, considerations in depth of how methodological flaws in the trial could lead to Type II errors.

Becker and Greig [1] and Becker [3] questioned whether or not drug development and CTs failed the drugs they tested because methodological deficiencies increased the probability of Type II errors. The low rates of attention we found to methodological issues that could invalidate drug development investigations, such as unreliability that leads to variance, reduced power, large numbers of subjects to meet power requirements, large numbers of sites to provide subjects, heterogeneous samples, inadequate monitoring and re-training of site personnel during studies, and so forth, sustain the concern that current AD CT methods and practices may lead to rejection of compounds that could be efficacious in AD or indicative of mechanisms of drug action efficacious in AD. Based on this added evidence we call for further research into AD clinical pharmacological methods and practices with the aims to make AD drug developments more effective, to increase confidence warranted in reported outcomes, to provide researchers improved tools for use in areas of special concern, such as outcome measure reliabilities and controlling sample sizes, and to transfer to

practitioners the methods, procedures, practices, protocols and other conditions of use of a drug shown in CTs as conditions for drug effectiveness. Readers need increased confidence in the validity of CT reports; however, to meet the clinical practice aims of medical research, improvements in health care, CTs need successful transfer to clinicians of information on how to use drugs effectively in patient care.

Becker and Greig [1] proposed, based on the available literature, that human error and limitations inherent in rating scales and neuropsychological tests precluded these tools from ever providing adequately precise and accurate information for CT testing of drugs and patient management with approved drugs. They suggested specific tightening up of current CT practices, further research into development of protocols to govern applications of methods and practices in research and the clinic, modification of intention-to-treat designs to allow CT investigators to change real world conditions that undermine demonstrations of effectiveness of AD candidate drugs and effective uses of approved drugs with patients, and accelerated development of quantitative biomarkers able to be used as surrogate outcomes in CTs and clinical practices. Imaging and biochemical monitoring currently offer specific advantages to AD researchers and practitioners, for example, exclusions of non-AD dementias and confirmation of AD [11]. These biomarkers can not currently serve as surrogate outcomes; yet, the increase we find in use of biomarkers as secondary outcome measures with drugs under development, 60% of studies sampled (Table II: 4. Query 37), and Phase II biomarker-based justifications for proceeding to Phase III [123] are in most proposals not accompanied by considerations of the validity of the proposed use (Table II: 4. Query 38). Advances proposed in methodologies, such as the use of the NTP battery for increased sensitivity to cognitive impairments, are not practical for clinicians who must depend on cognitive, behavioral, and global assessments to manage dosing and select among drug alternatives available to them when patients appear not to benefit from a current therapy.

As we earlier proposed, real improvements to CT validity and effective investigational information transfer to clinicians may await qualification of AD biomarkers as surrogate outcomes and practical guidance and community resources to support their use in AD and pre-AD patient care. One barrier to increased orientation of CTs towards biomarkers as outcome variables is a possible conflation of concepts of disease and patients' clinical status. A drug acts at molecular targets. Molecular targeting recommends biomarkers of molecular pathologies directly linked to clinical illness as preferred measures of outcome in CTs and for management in the clinic. Clinical benefits to patients and safety remain crucial as goals of treatment; however, as with many chronic diseases, prevention of early progression to AD and its disabilities may depend solely on management of biomarkers shown in long-term studies to be linked to later clinical illness. Given the evidence of pre-clinical progression of pathologies such as abnormalities in amyloid metabolism, we expect that early AD interventions will be managed based on biomarkers and not clinical assessments. Clinical assessments will test biomarkers for efficacy predicting long-term outcomes from molecular disease; drugs, through activities on systems monitored by biomarkers, will prevent clinical illness, disability, and handicap.

Our search may have missed publications and we thereby overestimate rates of failed publications. Our queries and methods are not tested and many aspects could be open to appropriate criticisms and refinements. On the other hand, we doubt that the full magnitude of inattention we found can be questioned and find support from our study for the results in already published reports that we used to develop our queries (see Tables II: 1-6). We find nothing in our current study that would undermine already existing concerns that methodological deficiencies and inadequately controlled procedural practices may be interfering with effective AD drug development and effective uses of AD drugs with

patients in the clinic. Given the low rates of success with AD drug candidates we repeat our earlier proposals that priority be given to research into AD clinical pharmacological methods, qualification of biomarkers as quantitative surrogate outcome measures, and redefinitions of CTs as providing both evidence of efficacy and of the conditions of use required in clinical practice to achieve efficacy.

Becker and Greig [1] questioned whether symptoms and adverse events associated with other diseases and medications could be mistaken by investigators as AD drug associated and account for failures to raise doses of AD drug to within prescribed ranges. We find high rates of other diagnoses being present and of non-AD medications prescribed in the studies we reviewed (Table II: 3. Queries 23 and 24). We do not find a clear association of these factors with failures to achieve target dosing (Table II: 2. Query 18) or with excessively high reported rates of discontinuation from drug due to adverse events (Table II: 2. Query 20). On the other hand, a proper investigation of this relationship is impossible due to the sporadic reporting in these areas of query.

The IOM report on PTSD research [6] claimed that many of their criticisms could be accounted for by earlier PTSD studies not providing the quality CTs currently available. If one assumes that roughly for our categories the AD approved drugs tap earlier AD research practices, AD failed the next period, MCI the early 21<sup>st</sup> century, and drugs under development current practices we see no trends on the majority of queries that would suggest increasing methodological soundness over time. We take this as suggestive that the IOM claim of improving research practices with time may not apply for AD and, given the references to failed methodologies in depression studies we reviewed as background, may be questioned for PTSD without direct evidence for that diagnosis to the contrary.

Specific attention may be required to the use of reliability coefficients to justify confidence in assessment methods. For example, Kang et al. [124] accepted test-retest correlations of 0.7 and 0.81 as validations of their interview methods for estimating cognitive status; yet, Becker and Markwell [21] found test-retest reliability coefficients ranging from 0.85 to 0.95 associated with levels of random measurement error greater than the average expected drug effects or disease decline over one year. Outcome measures must meet the levels required by the study for specific aspects of reliability, that is, precision, accuracy, sensitivity, specificity and so forth. The validity of indicators of reliability is too often neglected as part of research studies or proposals (Table II: 4. Query 33). We interpret the evidence from Becker and Markwell [21] for hidden random measurement error effects in data and from Engelhardt et al. [15] and others for widespread rater inaccuracies as requiring more than intra- and inter-rater reliabilities to ground validity for an investigation.

Our impression from our own work and the reports of others is that, although major steps away from current concerns with methodological reliability and validity await further research, much could be done immediately by greater attention to clinical pharmacological methods and practices. There is evidence that in both commercial developments and academic studies, as exemplified in recent MCI studies, before rushing to Phase III little or no attention is given to traditional Phase II aims to identify an effective dose and dosing range for a drug. For example, in the Feldman et al. [108] study of rivastigmine for MCI, it is not possible to decide whether the maximum 6 mg. dose achieved provides a fair test of the ChEI in MCI or under-dosing of subjects. As we earlier proposed for ChEIs [125] and as we demonstrated for metrifonate [19], ChEI may exhibit an inverted U shaped dose-response curve both within a population and within the range from normal to AD populations. Since there is most probably less cholinergic deficit in pre-AD than in AD, doses of ChEIs lower than those needed in AD may be indicated for MCI. Without Phase II preparations for the MCI studies the dosing issue can not be settled based on

pharmacological evidence leaving unanswered questions of possible Type II error interpretations due to mistaken dosing in current ChEI MCI studies.

Although studies of unreliability in outcomes of CTs dependent on rating scales are sparse we interpret a trend in investigations that supports very difficult, if not impossible, problems to overcome. Evidence of imprecision, inaccuracy, bias, lack of skills in raters especially in large studies, accrues to convince us to take seriously that human errors and probably limitations in rating scales, in neuropsychological tests, and in the sensitivity of clinical symptoms to neuropathological losses in pre-AD, preclude AD research from continuing its dependence on rated patient outcomes. In our view, current evidence supports that, as a disease, AD first expresses itself in pre-symptomatic neuropathological losses that are irreversible. This justifies the current emphasis on early diagnosis and early treatment but also undermines our dependence on rated symptomatology because research and interventions in CTs and the clinic will want to occur while persons are still asymptomatic or so slightly impaired as to be indistinguishable from many normal aged persons. Consequently we call for research into biomarkers and their developments as surrogate outcome to be given highest priority.

Many of the problems plaguing AD research currently—large sample sizes, multiple sites, sites without raters skilled in AD evaluation, problems of monitoring sites, low power and consequent risks of Type II errors—we find secondary to human error intruding into AD assessments. To reduce these error interferences we suggest a two phase drug clinical development model. Biomarkers, as surrogate markers of neuropathological changes leading to AD and its clinical progression, will become the outcome measures for CTs. Thal et al. [11], when they demonstrate the greatly reduced numbers of subjects needed to achieve required CT power, witness to the improvements in accuracy and precision achievable with carefully conducted biomarker studies. In a separate phase of research we propose investigators qualify biomarkers as surrogate markers for patient benefits. These latter activities can be carried out in AD specialized centers with well-experienced staff trained in AD assessment and evaluation and practiced regularly in these skills. With a range of the measures we have identified—ongoing training, qualification of raters, use only of highly skilled and experienced raters, use of means of three data points rather than single data points, long-term outcomes, such as onset of AD and so forth—we anticipate small but effective studies linking biomarker status to clinical benefits for patients [1]. Larger studies with community sites can develop and test protocols able to provide the reliabilities for biomarkers needed for efficient CTs and practitioners' uses with patients. Smaller studies qualify these biomarkers as surrogates and thus as acceptable outcomes for regulatory authorities.

We interpret our current study of a sample of AD drug outcome reports as adding to Aisen's, Vellas', others, and our own concerns with too many centers, too big variations among centers, difficulties finding homogeneous groups and managing large numbers of subjects in a CT, and other potential interferences with AD drug developments. We also find additional support for the proposal by Becker [3] and Becker and Greig [1] that problems of variance and its consequences can not be effectively overcome so long as CTs remain dependent on human rated outcome measures. We conclude that the summary by Thal et al. [11] of estimated sample sizes needed when biomarkers are used, combined with our and others evidence supporting how human error intrudes into effective AD drug development and evidence of those intrusions in this report, supports giving high priority to research into establishing biomarkers as quantitative and reliable indicators of long-term efficacy in AD CTs.



## Acknowledgments

The authors (REB and NHG) are supported in part by the Intramural Research Program of the National Institute on Aging, National Institutes of Health (NIH).

## References

1. Becker R, Greig NH. Alzheimer's disease drug development in 2008 and beyond: problems and opportunities. *Curr Alz Res.* 2008 April. In press.
2. Giacobini E, Becker RE. One hundred years after the discovery of Alzheimer's disease. A turning point for therapy? *Journal of Alzheimer's Disease.* 2007; 12:37–52.
3. Becker RE. Lessons from Darwin: 21st Century designs for clinical trials. *Curr Alz Res.* 2007; 4:458–467.
4. Dubois B, Feldman HH, Jacova C, DeKosky ST, Barberger-Gateau P, Cummings J, Delacourte A, Galasko D, Gauthier S, Jicha G, Meguro K, O'Brien J, Pasquier F, Pober P, Rossor M, Salloway S, Stern Y, Visser PJ, Scheltens P. Research criteria for the diagnosis of Alzheimer's disease: revising the NINCDS-ADRDA criteria. *Lancet Neurol.* 2007; 6:734–746. [PubMed: 17616482]
5. Pangalos MN, Schechter LE, Hurko O. Drug development for CNS disorders: strategies for balancing risk and reducing attrition. *Nature Reviews: Drug Discovery.* 2007; 6:521–532.
6. Institute of Medicine, Committee on Treatment of Post Traumatic Stress Disorder. *Treatment of Post Traumatic Stress Disorder, An Assessment of the Evidence.* Washington (DC): The National Academies Press; 2007.
7. American Psychiatric Association (APA), Practice guidelines. Acute stress disorder and posttraumatic stress disorder. *Amer J Psychiatry.* 2004; 161 Supplement
8. European Medicines Agency. EMEA/CHMP/EWP Workshop. Slowing the progression of neurodegenerative diseases: medicinal products clinical development. London: European Medicines Agency; 2006. Doc Ref. EMEA/512562/2006
9. Mani RB. The evaluation of disease modifying therapies in Alzheimer's disease: a regulatory viewpoint. *Stat Med.* 2004; 23:305–314. [PubMed: 14716731]
10. Pincus T, Stein CM. Why randomized controlled clinical trials do not depict accurately long-term outcomes in rheumatoid arthritis: some explanations and suggestions for future studies. *Clin Exp Rheumatol.* 1997; 15:S27–S38. [PubMed: 9266130]
11. Thal LJ, Kantarci K, Reiman EM, Klunk WE, Weiner MW, Zetterberg H, Galasko D, Praticò D, Griffin S, Schenk D, Siemers E. The role of biomarkers in clinical trials for Alzheimer disease. *Alzheimer Dis Assoc Disord.* 2006; 20:6–15. [PubMed: 16493230]
12. Butcher J. The hunt for drugs to modify Alzheimer's disease. *Lancet Neurology.* 2007; 6:1038–1039. [PubMed: 18031701]
13. Dworkin RH, Katz J, Gitlin MJ. Placebo response in clinical trials of depression and its implications for research on chronic neuropathic pain. *Neurology.* 2005; 65:S17–S19.
14. Becker RE. Lessons from Darwin: Evolutionary biology implications for Alzheimer's disease research and patient care. *Curr Alz Res.* 2005; 2:319–326.
15. Engelhardt N, Feiger AD, Cogger KO, Sikich D, DeBrotta DJ, Lipsitz JD, Kobak KA, Evans KR, Potter WZ. 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]
16. Easterbrook PJ, Berlin JA, Gopalan R, Matthews DR. Publication bias in clinical research. *Lancet.* 1991; 337:867–72. [PubMed: 1672966]
17. Frampton M, Harvey RJ, Kirchner V. Propentofyllin for dementia. *Cochrane Data Base of Systematic Reviews.* 2003; 2:CD002853.
18. Chalmers I. Underreporting research is scientific misconduct. *JAMA.* 1990; 263:1405–1408. [PubMed: 2304220]
19. Becker RE, Colliver JA, Markwell SJ, Moriearty PL, Unni LK, Vicari S, Double-Blind A. Placebo-Controlled Study of Metrifonate, an Acetylcholinesterase Inhibitor, for Alzheimer's Disease. *Alzheimer's Disease and Associated Disorders.* 1996; 10:124–131.

20. Becker RE, Colliver JA, Markwell SJ, Moriearty PL, Unni LK, Varney A, Vicari S. A Six Month Study of the Effects of Metrifonate on Cognitive Decline in Alzheimer's Disease. *Alzheimer's Disease and Associated Disorders*. 1998; 12:54–57.
21. 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.
22. Perkins DO, Wyatt RJ, Bartko JJ. Penny-wise and pound-foolish: the impact of measurement error on sample size requirements in clinical trials. *Biol Psychiatry*. 2000; 47:762–766. [PubMed: 10773186]
23. Lachin JM. The role of measurement reliability in clinical trials. *Clin Trials*. 2004; 1:553–566. [PubMed: 16279296]
24. Woods SW, Gueorguieva RV, Baker CB, Makuch RW. Control group bias in randomized atypical antipsychotic medication trials for schizophrenia. *Arch Gen Psychiatry*. 2005; 62:961–970. [PubMed: 16143728]
25. Kobak, KA.; DeBrot, DJ.; Engelhardt, N.; Williams, JBW. Site vs. centralized raters in a clinical depression trial, National Institute of Mental Health, New Clinical Drug Evaluation Unit. 46th Annual Meeting; Boca Raton, Fl. 2006, June;
26. Varma AR, Snowden JS, Lloyd JJ, Talbot PR, Mann DM, Neary D. Evaluation of the NINCDS-ADRDA criteria in the differentiation of Alzheimer's disease and frontotemporal dementia. *J Neurol Neurosurg Psychiatry*. 1999; 66:184–8. [PubMed: 10071097]
27. 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]
28. Khan A, Leventhal RM, Khan SR, Brown WA. Severity of depression and response to antidepressants and placebo: analysis of the Food and Drug Administration database. *J Clin Psychopharmacol*. 2002; 22:40–45. [PubMed: 11799341]
29. Thase ME. Studying new antidepressants: if there were a light at the end of the tunnel could we see it? *J Clin Psychiatry*. 2002; 63 2:24–28. [PubMed: 15453011]
30. Leon AC, Marzuk PM, Portera L. More reliable outcome measures can reduce sample size requirements. *Arch Gen Psychiatry*. 1995; 52:867–871. [PubMed: 7575107]
31. Harrison J, Minassian SL, Jenkins L, Black RS, Koller M, Grundman M. A neuropsychological test battery for use in Alzheimer disease clinical trials. 2007; 64:1323–1329.
32. Muller MJ, Szegedi A. Effects of interrater reliability of psychopathologic assessments on power and sample size calculations in clinical trials. *J Clin Psychopharmacol*. 2002; 22:318–325. [PubMed: 12006903]
33. 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]
34. Hamilton M. A rating scale for depression. *J Neurology Neurosurg & Psychiatry*. 1960; 23:56–62.
35. Demitrack MA, Faries D, Herrera JM, DeBrot, D, Potter WZ. The problem of measurement error in multisite clinical trials. *Psychopharmacol Bull*. 1998; 34:19–24. [PubMed: 9564194]
36. Kobak KA, Lipsitz JD, Feiger A. Development of a standardized training program for the Hamilton Depression Scale using internet-based technologies: results from a pilot study. *J Psychiatr Res*. 2003; 37:509–15. [PubMed: 14563382]
37. Kobak KA, Lipsitz JD, Williams JB, Engelhardt N, Bellew KN. A new approach to rater training and certification in a multicenter clinical trial. *J Clin Psychopharmacol*. 2005; 25:407–412. [PubMed: 16160614]
38. Targum SD. Evaluating rater competency for CNS clinical trials. *J Clin Psychopharmacol*. 2006; 26:308–310. [PubMed: 16702896]
39. Rickels K, Robinson DS. Why do clinical trials fail? *J Clin Psychopharmacol*. 2007; 27:420–421. [PubMed: 17632242]
40. Kobak KA, Kane JM, Thase TE, Nierenberg AA. Reply to comments by Drs Rickels and Robinson. *J Clin Psychopharmacol*. 2007; 27:421.
41. Cummings JL. Clinical evaluation as a biomarker for Alzheimer's disease. *J Alz Dis*. 2005; 8:327–337.

42. Winblad, B.; Giacobini, E.; Froelich, L.; Bruinsma, G.; Walters, E.; Friedhoff, L. Double-blind placebo-controlled evaluation of the safety and efficacy of phenserine tartrate for the treatment of mild to moderate Alzheimer disease. 9th International Geneva/Springfield Symposium on Advances in Alzheimer Therapy; April (2006); p. 19-22. Abstract 11A
43. Qureshi NN, Hatcher J, Chaturvedi N, Jafar TH, Hypertension Research Group. Effect of general practitioner education on adherence to antihypertensive drugs: cluster randomized controlled trial. *BMJ*. 2007; 335:1030–1037. [PubMed: 17991935]
44. Dybicz SB, Keohane DJ, Erwin WG, McRae T, Shah SN. Patterns of cholinesterase-inhibitor use in the nursing home setting: a retrospective analysis. *Am J Geriatr Psychother*. 2006; 4:154–160.
45. Schmidt R, Lechner A, Petrovic K. Rivastigmine in outpatient services: experience of 114 neurologists in Austria. *Int Clin Psychopharmacol*. 2002; 17:81–85. [PubMed: 11890189]
46. Bramley TJ, Gerbino PP, Nightengale BS, Frech-Tamas F. Relationship of blood pressure control to adherence with antihypertensive monotherapy in 13 managed care organizations. *J Manag Care Pharm*. 2006; 12:239–245. [PubMed: 16623608]
47. Bellilli G, Lucchi E, Minicuci N, Rozzini L, Bianchetti A, Padovani A, Trabucchi M. Results of a multi-level therapeutic approach for Alzheimer's disease subjects in the "real world" (CRONOS project): a 36-week follow-up study. *Aging Clin Exp Res*. 2005; 17:54–61. [PubMed: 15847123]
48. Hallak M, Giacobini E. Physostigmine, tacrine, and metrifonate: the effect of multiple doses on acetylcholine metabolism in rat brain. *Neuropharmacology*. 1989; 28:190–206.
49. Hallak M, Giacobini E. A comparison of the effects of two inhibitors on brain cholinesterase. *Neuropharmacology*. 1987; 26:521–530. [PubMed: 3601008]
50. Mori F, Cuadra G, Giacobini E. Metrifonate effects on acetylcholine and biogenic amines in rat cortex. *Neurochemical Research*. 1995; 20:1081–1088. [PubMed: 8570013]
51. Becker RE, Giacobini E. Mechanisms of cholinesterase inhibition in senile dementia of the Alzheimer type: clinical, pharmacological, and therapeutic aspects. *Drug Dev Res*. 1988; 12:163–195.
52. Becker R, Colliver J, Elble R, Feldman E, Giacobini E, Kumar V, Markwell S, Moriearty P, Parks R, Shillcutt S, Unni L, Vicari S, Womack C, Zec R. Effects of Metifonate, A Long Acting Cholinesterase Inhibitor in Alzheimer Disease: Report of an Open Trial. *Drug Dev Res*. 1990; 9:425–434.
53. Carroll BJ. Letter to Editor. *Lancet*. 2006; 367:1235–1236. [PubMed: 16631869]
54. Kelloff GJ, Bast RC, Coffey DS, D'Amico AV, Kerbel RS, Park JW, Ruddon RW, Rustin GJS, Schilsky RL, Sigman CC, Woude GFV. Biomarkers, surrogate end points, and the acceleration of drug development for cancer prevention and treatment: an update. *Clinical Cancer Research*. 2004; 10:3881–3884. [PubMed: 15173097]
55. Raschetti, R.; Albanese, E.; Vanacore, N.; Maggini, M. Cholinesterase inhibitors in mild cognitive impairment: a systematic review of randomized trials; *Plos Medicine*. 2007. p. 1818-1828. [www.plosmedicine.org](http://www.plosmedicine.org)
56. Williams JBW. A structured interview guide for the Hamilton Depression Rating Scale. *Arch Gen Psychiatry*. 1988; 45:742–747. [PubMed: 3395203]
57. Petersen RC. MCI treatment trials: failure or not? *Lancet Neurology*. 2007; 6:473–475. [PubMed: 17509475]
58. Beers, Y. *Introduction to the Theory of Error*. Palo Alto: Addison-Wesley Publishing Co. Inc.; 1959.
59. Ferris SH. Are placebo controls still placebo controls? *Neurobiology of Aging*. 2006; 27 S1:S29.
60. Schneider LS. Relevance of clinical trials: can new strategies offer more? *Neurobiology of Aging*. 2006; 27 S1:S76.
61. Foster, N. *Imaging Advances for Early Detection and New Drug Assessment in Alzheimer's Disease*. Clinical Neurosciences Center; University of Utah: Nov 28. 2007
62. Weiner, M. Appendix. Introductory section background and significance of the overall proposal. Nov 28. 2007 [Hppt:www.loni.ucla.edu/ADNI/pdf/ADNI\\_Introduction.pdf](http://www.loni.ucla.edu/ADNI/pdf/ADNI_Introduction.pdf). Viewed
63. Lopez OL, Swihart AA, Becker JT, Reinmuth OM, Reynolds CF, Rezek DL, Daly FL. Reliability of NINCDS-ADRDA clinical criteria for the diagnosis of Alzheimer's disease. *Neurology*. 1990; 40:1517–22. [PubMed: 2215942]

64. Temple R. Are Surrogate Markers Adequate to Assess Cardiovascular Disease Drugs? *JAMA*. 1999; 282:790–795. [PubMed: 10463719]
65. Katz R. FDA Evidentiary Standards for Drug Development and Approval. *NeuroRx*. 2004; 1:307–316. [PubMed: 15717032]
66. Nordic Council on Medicines. Good clinical trial practice: Nordic Guidelines, NLN Publication No 28. Uppsala, Sweden: 1989.
67. Food and Drug Administration. FDA. US Department of Health and Human Services; Rockville, MD, USA: 1989. Guideline for the format and content of the clinical and statistical sections of new drug applications.
68. Kantarci K, Jack CR. Neuroimaging in Alzheimer disease: an evidence-based review. *Neuroimaging Clin N Am*. 2003; 13:197–209. [PubMed: 13677801]
69. Jack CR Jr, Weigand SD, Shiung MM, Przybelski SA, O'Brien PC, Gunter JL, Knopman DS, Boeve BF, Smith GE, Petersen RC. Atrophy rates accelerate in amnesic mild cognitive impairment. *Neurology*. Nov 21. 2007 Epub ahead of print.
70. Jack CR Jr, Shiung MM, Weigand SD, O'Brien PC, Gunter JL, Boeve BF, Knopman DS, Smith GE, Ivnik RJ, Tangalos EG, Petersen RC. Brain atrophy rates predict subsequent clinical conversion in normal elderly and amnesic MCI. *Neurology*. 2005; 65:1227–1231. [PubMed: 16247049]
71. Jack CR Jr, Dickson DW, Parisi JE, Xu YC, Cha RH, O'Brien PC, Edland SD, Smith GE, Boeve BF, Tangalos EG, Kokmen E, Petersen RC. Antemortem MRI findings correlate with hippocampal neuropathology in typical aging and dementia. *Neurology*. 2002; 58:750–757. [PubMed: 11889239]
72. Jack CR Jr, Petersen RC, Xu Y, O'Brien PC, Smith GE, Ivnik RJ, Boeve BF, Tangalos EG, Kokmen E. Rates of hippocampal atrophy correlate with change in clinical status in aging and AD. *Neurology*. 2000; 55:484–89. [PubMed: 10953178]
73. den Heijer T, Geerlings MI, Hoebeek FE, Hofman A, Koudstaal PJ, Breteler MM. Use of hippocampal and amygdalar volumes on magnetic resonance imaging to predict dementia in cognitively intact elderly people. *Arch Gen Psychiatry*. 2006; 63:57–62. [PubMed: 16389197]
74. Grundman M, Jack CR, Petersen RC, Kim HT, Taylor C, Davtarian M, Weiner MF, DeCarli C, DeKosky ST, van Dyck C, Darvesh S, Yaffe K, Kaye J, Ferris SH, Thomas RG, Thal LJ. The Alzheimer's Disease Cooperative Study, Hippocampal volume is associated with memory but not nonmemory cognitive performance in patients with mild cognitive impairment. *J Mol Neurosci*. 2003; 20:241–248. [PubMed: 14501003]
75. Nordberg A. PET imaging of amyloid in Alzheimer's disease. *Lancet Neurol*. 2004; 3:519–527. [PubMed: 15324720]
76. Frisoni GB, Scheltens P, Galluzzi S, Nobili FM, Fox NC, Robert PH, Soininen H, Wahlund LO, Waldemar G, Salmon E. Neuroimaging tools to rate regional atrophy, subcortical cerebrovascular disease, and regional cerebral blood flow and metabolism: consensus paper of the EADC. *J Neurol Neurosurg Psychiatry*. 2003; 74:1371–1381. [PubMed: 14570828]
77. Frisoni GB, Testa C, Sabatoli F, Beltramello A, Soininen H, Laakso MP. Structural correlates of early and late onset Alzheimer's disease: voxel based morphometric study. *J Neurol Neurosurg Psychiatry*. 2005; 76:112–114. [PubMed: 15608008]
78. Fox NC, Schott JM. Imaging cerebral atrophy: normal ageing to Alzheimer's disease. *Lancet*. 2004; 363:392–394. [PubMed: 15074306]
79. Minoshima S, Giordani B, Berent S, Frey KA, Foster NL, Kuhl DE. Metabolic reduction in the posterior cingulate cortex in very early Alzheimer's disease. *Ann Neuro*. 1997; 42:85–94.
80. Chan D, Janssen JC, Whitwell JL, Watt HC, Jenkins R, Frost C, Rossor MN, Fox NC. Change in rates of cerebral atrophy over time in early-onset Alzheimer's disease: longitudinal MRI study. *Lancet*. 2003; 362:1121–1122. [PubMed: 14550701]
81. Scheltens P, Fox N, Barkhof F, De Carli C. Structural magnetic resonance imaging in the practical assessment of dementia: beyond exclusion. *Lancet Neurol*. 2002; 1:13–21. [PubMed: 12849541]
82. Fox NC, Crum WR, Scahill RI, Stevens JM, Janssen JC, Rossor MN. Imaging of onset and progression of Alzheimer's disease with voxel-compression mapping of serial magnetic resonance images. *Lancet*. 2001; 358:201–205. [PubMed: 11476837]

83. Edison P, Archer HA, Hinz R, Hammers A, Pavese N, Tai YF, Hotton G, Cutler D, Fox N, Kennedy A, Rossor M, Brooks DJ. Amyloid, hypometabolism, and cognition in Alzheimer disease: an [11C]PIB and [18F]FDG PET study. *Neurology*. 2007; 68:501–508. [PubMed: 17065593]
84. Foster NL, Heidebrink JL, Clark CM, Jagust WJ, Arnold SE, Barbas NR, DeCarli CS, Turner RS, Koeppe RA, Higdon R, Minoshima S. FDG-PET improves accuracy in distinguishing frontotemporal dementia and Alzheimer's disease. *Brain*. 2007; 130:2616–2635. [PubMed: 17704526]
85. Jagust W, Thisted R, Devous MD Sr, Van Heertum R, Mayberg H, Jobst K, Smith AD, Borys N. SPECT perfusion imaging in the diagnosis of Alzheimer's disease: a clinical-pathologic study. *Neurology*. 2001; 56:950–956. [PubMed: 11294935]
86. Klunk WE, Engler H, Nordberg A, Wang Y, Blomqvist G, Holt DP, Bergström M, Savitcheva I, Huang GF, Estrada S, Ausén B, Debnath ML, Barletta J, Price JC, Sandell J, Lopresti BJ, Wall A, Koivisto P, Antoni G, Mathis CA, Långström B. Imaging brain amyloid in Alzheimer's disease with Pittsburgh Compound-B. *Ann Neurol*. 2004; 55:306–319. [PubMed: 14991808]
87. Alexander GE, Chen K, Pietrini P, Rapoport SI, Reiman EM. Longitudinal PET Evaluation of Cerebral Metabolic Decline in Dementia: A Potential Outcome Measure in Alzheimer's Disease Treatment Studies. *Am J Psychiatry*. 2002; 159:738–745. [PubMed: 11986126]
88. Engler H, Forsberg A, Almkvist O, Blomqvist G, Larsson E, Savitcheva I, Wall A, Ringheim A, Langstrom B, Nordberg A. Two-year followup of amyloid deposition in patients with Alzheimer's disease. *Brain*. 2006; 129:2856–2866. [PubMed: 16854944]
89. Wolf H, Jelic V, Gertz HJ, Nordberg A, Julin P, Wahlund LO. A critical discussion of the role of neuroimaging in mild cognitive impairment. *Acta Neurologica Scandinavica*. 2003; 107:52–76.
90. Zhong Z, Ewers M, Teipel S, Burger K, Wallin A, Blennow K, He P, McAllister C, Hampel H, Shen Y. Levels of gamma secretase I cerebrospinal fluid as a predictor of risk in mild cognitive impairment. *Arch Gen Psychiatry*. 2007; 64:718–726. [PubMed: 17548753]
91. Buerger K, Ewers M, Pirttila T, Zinkowski R, Alafuzoff I, Teipel SJ, DeBernardis J, Kerkman D, McCulloch C, Soininen H, Hampel H. CSF phosphorylated tau protein correlates with neurofibrillary pathology in Alzheimer's disease. *Brain*. 2006; 129:3035–3041. [PubMed: 17012293]
92. Stefani A, Martorana A, Bernardini S, Panella M, Mercati F, Orlacchio A, Pierantozzi M. CSF markers in Alzheimer disease patients are not related to the different degree of cognitive impairment. *J Neurol Sci*. 2006; 251:124–128. [PubMed: 17097109]
93. Huey ED, Mirza N, Putnam KT, Soares H, Csako G, Levy JA, Copenhaver B, Cohen RM, Sunderland T. Stability of CSF beta-amyloid(1-42) and tau levels by APOE genotype in Alzheimer patients. *Dement Geriatr Cogn Disord*. 2006; 22:48–53. [PubMed: 16682793]
94. Tapiola T, Pirttila T, Mikkonen M, Mehta PD, Alafuzoff I, Koivisto K, Soininen H. Three-year follow-up of cerebrospinal fluid tau, beta-amyloid 42 and 40 concentrations in Alzheimer's disease. *Neurosci Lett*. 2000; 18:119–122. [PubMed: 10686392]
95. Hansson O, Zetterberg H, Buchhave P, Andreasson U, Londos E, Minthon L, Blennow K. Prediction of Alzheimer's disease using the CSF Abeta42/Abeta40 ration in patients with mild cognitive impairments. *Dement Geriatr Cogn Disord*. 2007; 23:316–320. [PubMed: 17374949]
96. Yan Y, Wang C. Abeta40 protects non-toxic Abeta42 monomer from aggregation. *J Mol Biol*. 2007; 369:909–916. [PubMed: 17481654]
97. Kim J, Onstead L, Randle S, Price R, Smithson L, Zwizinski C, Dickson DW, Golde T, McGowan E. Abeta40 inhibits amyloid deposition in vivo. *J Neurosci*. 2007; 17:627–633. [PubMed: 17234594]
98. Lesko LJ, Atkinson AJ. Use of biomarkers and surrogate endpoints in drug development and regulatory decision making: criteria, validation, strategies. *Annual Review of Pharmacology and Toxicology*. 2001; 41:347–366.
99. Granger CB, McMurray JJV. Using measures of disease progression to determine therapeutic effect: a Siren's song. *J Am Coll Cardiol*. 2006; 48:434–437. [PubMed: 16875965]
100. Molenberghs, G.; Burzykowski, T.; Alonso, A.; Buyse, M. Statistical methods in medical research. Vol. 13. 2004. p. 177-206.

101. Patsy BM, Weiss NS, Furberg CD, Koepsell TD, Siscovick DS, Rosendaal FR, Smith NL, Heckbert SR, Kaplan RC, Lin D, Fleming TR, Wagner EH. Surrogate end points, health outcomes, and the drug-approval process for the treatment of risk factors in cardiovascular disease. *JAMA*. 1999; 282:786–790. [PubMed: 10463718]
102. Kralawish J. Alzheimer's disease—clinical trials and logic of clinical purpose. *N Engl J Med*. 2006; 355:1604–1605. [PubMed: 17035654]
103. Kent DM, Hayward RA. Limitations of applying summary results of clinical trials to individual patients. *JAMA*. 2006; 298:1209–1212. [PubMed: 17848656]
104. Califf RM. Simple principles of clinical trials remain powerful. *JAMA*. 2005; 293:489–491. [PubMed: 15671435]
105. Tunis SR, Stryer DB, Clancy CM. Practical clinical trials: increasing the value of clinical research for decision making in clinical and health policy. *JAMA*. 2003; 290:1624–1632. [PubMed: 14506122]
106. Rich MW. From clinical trials to clinical practice. *JAMA*. 2002; 287:1321–1323. [PubMed: 11886324]
107. Becker RE. Focusing pharmaceutical research on patient care: using clinical trials to develop clinical decision rules. *Clin Drug Invest*. 2002; 22:269–278.
108. Feldman HH, Ferris S, Winblad B, Sfikas N, Mancione L, He Y, Tekin S, Burns A, Cummings J, del Ser T, Ogogozo-M JM, Sauer H, Scheltens P, Scarpini E, Herrmann N, Farlow M, Potkin S, Charles HC, Fox NC, Lane R. Effect of rivastigmine on delay to diagnosis of Alzheimer's disease from mild cognitive impairment: the InDDEx study. *Lancet Neurology*. 2007; 6:501–512. [PubMed: 17509485]
109. Jelic V. Clinical trials in MCI and what we have learned. *Neurobiology of Aging*. 2006 27S1:S47.
110. Brys M, Pirraglia E, Rich K, Mosconi L, Switalski R, Glodzik-Sobanska L, De Santi S, Zinkowski R, Mehta P, Praticao D, Saint Louis LA, Wallin A, Blennow K, Leon MJ. Prediction and longitudinal study of CSF biomarkers in mild cognitive impairment. *Neurobiol Aging*. 2007 e-pub ahead of print.
111. Blennow K, Hampel H. CSF markers for incipient Alzheimer's disease. *Lancet Neurol*. 2003; 2:605–613. [PubMed: 14505582]
112. Simonsen AH, McGuire J, Podust VN, Hagnelius NO, Nilsson TK, Kapaki E, Vassilopoulos D, Waldemar G. A Novel Panel of Cerebrospinal Fluid Biomarkers for the Differential Diagnosis of Alzheimer's Disease versus Normal Aging and Frontotemporal Dementia. *Dement Geriatr Cogn Disord*. 2007; 24:434–440. [PubMed: 17971664]
113. Simonsen AH, McGuire J, Podust VN, Davies H, Minthon L, Skoog I, Andreassen N, Wallin A, Waldemar G, Blennow K. Identification of a novel panel of cerebrospinal fluid biomarkers for Alzheimer's Disease. *Neurobiol Aging*. 2007 Epub ahead of print.
114. Gauthier S, Touchon J. Mild cognitive impairment is not a clinical entity and should not be treated. *Arch Neurol*. 2005; 62:1164–1166. [PubMed: 16009780]
115. Visser PJ, Scheltens P, Verhey FR. Do MCI criteria in drug trials accurately identify subjects with predementia Alzheimer's disease? *J Neurol Neurosurg Psychiatry*. 2005; 76:1348–1354. [PubMed: 16170074]
116. Kwon MO, Herrling P. List of drugs in development for neurodegenerative diseases. *Neurodegenerative Dis*. 2005; 2:61–108.
117. Sano M, Ernesto C, Thomas RG, Klauber MR, Schafer K, Grundman M, Woodbury P, Growdon J, Cotman CW, Pfeiffer E, Schneider LS. A controlled trial of selegiline, alpha-tocopherol, or both as treatment for Alzheimer's disease. *N Eng J Med*. 1997; 336:1216–1222.
118. Marcusson J, Rother M, Kittner B, Rössner M, Smith BJ, Babic T, Folnegovic-Smalc V, Möller HJ, Labs KH. A 12-month, randomized, placebo-controlled trial of propentofylline (HWA 285) in patients with dementia according to DSM III-R. The European Propentofylline Study Group. *Dement Geriatr Cogn Disord*. 1997; 8:320–328. [PubMed: 9298634]
119. Mazza M, Capuano A, Bria P, Mazza S. Ginkgo bilboa and donepezil: a comparison in the treatment of Alzheimer's dementia in a randomized placebo-controlled double-blind study. *European Journal of Neurology*. 2006; 13:981–985. [PubMed: 16930364]

120. Moreno MDJM. Cognitive improvement in mild to moderate Alzheimer's dementia after treatment with the acetylcholine precursor choline alfoscerate: a multicenter, double-blind, randomized, placebo-controlled trial. *Clinical Therapeutics*. 2003; 25:178–193. [PubMed: 12637119]
121. Forest Laboratories. (Viewed Dec. 2007) MEM-MD-01  
[http://www.forestclinicaltrials.com/CTR/CTRController/CTRViewPdf?\\_file\\_id=scsr/SCSR\\_MEM-MD-01\\_final.pdf](http://www.forestclinicaltrials.com/CTR/CTRController/CTRViewPdf?_file_id=scsr/SCSR_MEM-MD-01_final.pdf)
122. Young JR. 'Superarchives' could hold all scholarly output. *The Chronicle of Higher Education*. 2002; 48:43.
123. Aisen PS, Gauthier S, Vellas B, Briand R, Saumier D, Laurin J, Garceau D. Alzhemed: a potential treatment for Alzheimer's disease. *Curr Alz Res*. 2007; 4:473–478.
124. Kang JH, Cook N, Manson J, Buring JE, Grodstein F. Low dose aspirin and cognitive function in the women's health study cognitive cohort. *BMJ*. 2007; 334:987–994. [PubMed: 17468120]
125. Becker, RE. Becker, R.; Giacobini, E. *Cholinergic Basis for Alzheimer Therapy*. Boston: Birkhauser; 1991. *Therapy of the cognitive deficit in Alzheimer's disease: the cholinergic system*.
126. Farlow M, Gracon SI, Hershey LA, Lewis KW, Sadowsky CH, Dolan-Ureno J. A 12-week, double-blind, placebo-controlled, parallel-group study of tacrine in patients with probable Alzheimer's disease. *JAMA*. 1992; 268:2523–2529. [PubMed: 1404819]
127. Wood PC, Castleden CM. A double-blind, placebo-controlled, multicenter study of tacrine for Alzheimer's disease. *Int J Geriatric Psychiatry*. 1994; 9:649–654.
128. Davis KL, Thal LJ, Gamzu ER, Davis CS, Woolson RF, Gracon SI, Drachman DA, Schneider LS, Whitehouse PJ, Hoover TM. The Tacrine Collaborative Study Group, Tacrine in patients with Alzheimer's disease: a double-blind placebo-controlled multicenter study. *N Engl J Med*. 1992; 327:1253–1259. [PubMed: 1406817]
129. Feldman H, Gauthier S, Hecker J, Vellas B, Subbiah P, Whalen E. Donepezil MSAD Study Investigators Group, A 24-week, randomized, double-blind study of donepezil in moderate to severe Alzheimer's disease. *Neurology*. 2001; 57:613–620. [PubMed: 11524468]
130. Winblad B, Kilander L, Eriksson S, Minthon L, Batsman S, Wetterholm AL, Jansson-Blixt C, Haglund A. for the Severe Alzheimer's Disease Study Group. Donepezil in patients with severe Alzheimer's disease: double-blind, parallel-group, placebo-controlled study. *Lancet*. 2006; 367:1057–1065. [PubMed: 16581404]
131. Winbald B, Engedal K, Soinenen H, Verhey F, Waldemar G, Wimo A, Wetterholm AL, Zhang R, Haglund A, Subbiah P, Donepezil Nordic Study Group. A 1-year, randomized, placebo-controlled study of donepezil in patients with mild to moderate AD. *Neurology*. 2001; 57:489–495. [PubMed: 11502918]
132. Ballard C, Margallo-Lana M, Juszcak E, Douglas S, Swann A, Thomas A, O'Brien J, Everatt A, Sadler S, Maddison C, Lee L, Bannister C, Elvish R, Jacoby R. Quetiapine and rivastigmine and cognitive decline in Alzheimer's disease: randomized double blind placebo controlled trial. *BMJ*. 2005; 330:874–878. [PubMed: 15722369]
133. Tariot PN, Farlow MR, Grossberg GT, Graham SM, McDonald S, Gergel I. for the Memantine Study Group. Memantine treatment in patients with moderate to severe Alzheimer disease already receiving donepezil: a randomized controlled trial. *JAMA*. 2004; 291:317–324. [PubMed: 14734594]
134. Reisberg B, Doody R, Stoffler A, Schmitt F, Ferris S, Mobius HJ. for the Memantine Study Group. Memantine in moderate-to-severe Alzheimer's disease. *N Eng J Med*. 2003; 348:1333–1341.
135. Mohr E, Nair NP, Sampson M, Murtha S, Belanger G, Pappas B, Mendis T. Treatment of Alzheimer's disease with sabeluzole: functional and structural correlates. *Clin Neuropharmacol*. 1997; 20:338–345. [PubMed: 9260731]
136. Potter A, Corwin J, Lang J, Piasecki M, Lenox R, Newhouse PA. Acute effects of the selective cholinergic channel activator (nicotinic agonist) ABT-418 in Alzheimer's disease. *Psychopharmacol*. 1999; 142:334–342.
137. Notice development discontinued viewed March 28, 2008, at  
<http://www.alzforum.org/drg/drc/detail.asp?id=75>,

138. Notice development discontinued viewed March 28, 2008 at  
[http://google.rsc.org/search?  
q=Alvamlene&site=Default&btnG.x=13&btnG.y=10&btnG=Search&output=xml\\_no\\_dtd&sort  
=date%3AD%3AL%3Ad1&ie=UTF-8&lr=&client=default\\_frontend&y=10&oe=&proxystylesheet=default\\_frontend  
d&x=11](http://google.rsc.org/search?q=Alvamlene&site=Default&btnG.x=13&btnG.y=10&btnG=Search&output=xml_no_dtd&sort=date%3AD%3AL%3Ad1&ie=UTF-8&lr=&client=default_frontend&y=10&oe=&proxystylesheet=default_frontend&x=11)
139. Cutler NR, Seifert RD, Schleman MM, Sramek JJ, Szyllayko OJ, Howard DR, Barchowsky A, Wardle TS, Brass EP. Acetylcholinesterase inhibition by zilfrosilone: pharmacokinetics and pharmacodynamics. *Clin Pharmacol Therapeutics*. 1995; 58:54–61.
140. Gaultieri F, Manetti D, Romanelli MN, Ghelardini C. Design and study of piracetam-like nootropics: controversial members of the problematic class of cognition-enhancing drugs. *Current Pharmaceutical Design*. 2002; 8:125–138. [PubMed: 11812254]
141. Cutler NR, Shrotriya RC, Sramek JJ, Veroff AE, Seifert RD, Reich LA, Hironaka DY. The use of the Computerized Neuropsychological Test Battery (CNTB) in an efficacy and safety trial of BMY 21,502 in Alzheimer's disease. *Ann N Y Acad Sci*. 1993; 695:332–336. [PubMed: 8239308]
142. Salloway S, Ferris S, Kluger A, Goldman R, Griesing T, Kumar D, Richardson S. for the Donepezil “401” Study Group, Efficacy of donepezil in mild cognitive impairment: a randomized placebo-controlled trial. *Neurology*. 2004; 63:651–657. [PubMed: 15326237]
143. Petersen RC, Thomas RG, Grundman M, Bennett D, Doody R, Ferris S, Galasko D, Jin S, Kaye J, Levey A, Pfeiffer E, Sano M, van Dyck CH, Thal LJ. Alzheimer's Disease Cooperative Study Group, Vitamin E and donepezil for the treatment of mild cognitive impairment. *N Engl J Med*. 2005; 352:2379–2388. [PubMed: 15829527]
144. Clinicalstudyresults.org. March 28. 2008  
[http://www.clinicalstudyresults.org/documents/company-study\\_96\\_1.pdf](http://www.clinicalstudyresults.org/documents/company-study_96_1.pdf) Viewed
145. Clinicalstudyresults.org. March 28. 2008  
[http://www.clinicalstudyresults.org/documents/company-study\\_96\\_2.pdf](http://www.clinicalstudyresults.org/documents/company-study_96_2.pdf) Viewed
146. ADAPT Research Group. Lyketos CG, Breitner JC, Green RC, Martin BK, Meinert C, Piantadosi S, Sabbagh M. Naproxen and celecoxib do not prevent AD in early results from a randomized controlled trial. *Neurology*. 2007; 68:1800–1808. [PubMed: 17460158]
147. ClinicalTrials.gov. March 28. 2008  
<http://clinicaltrials.gov/ct2/show/NCT00567060?term=piracetam&rank=1> Viewed
148. Aisen PS, Schafer KA, Grundman M, Pfeiffer E, Sano M, Davis KL, Farlow MR, Jin S, Thomas RG, Thal LJ, Alzheimer's Disease Cooperative Study. Effects of rofecoxib or naproxen vs placebo on Alzheimer disease progression: a randomized controlled trial. *JAMA*. 2003; 289:2819–2826. [PubMed: 12783912]
149. Resnick SM, Coker LH, Maki PM, Rapp SR, Espeland MA, Shumaker SA. The Women's Health Initiative Study of Cognitive Aging (WHISCA): a randomized clinical trial of the effects of hormone therapy on age-associated cognitive decline. *Clin Trials*. 2004; 1:440–450. [PubMed: 16279282]
150. ClinicalTrials.gov. March 28. 2008  
<http://clinicaltrials.gov/ct2/show/NCT00083590?term=HUPERZINE&rank=1> Viewed
151. ClinicalTrials.gov. March 28. 2008  
<http://clinicaltrials.gov/ct2/show/NCT00479557?term=acc-001&rank=1> Viewed
152. ClinicalTrials.gov. March 28. 2008  
<http://clinicaltrials.gov/ct2/show/NCT00141661?term=TTP+488&rank=2> Viewed
153. Galasko, DR.; Graff-Radford, N.; May, S.; Hendrix, S.; Cottrell, BA.; Sagi, SA.; Mather, G.; Laughlin, M.; Zavitz, KH.; Swabb, E.; Golde, TE.; Murphy, MP.; Koo, EH. *Alzheimer Dis Assoc Disord*. Vol. 21. 2007. p. 292-299.
154. Alvarez XA, Cacabelos R, Laredo M, Couceiro V, Sampedro C, Varela M, Corzo L, Fernandez-Novoa L, Vargas M, Aleixandre M, Linares C, Granizo E, Muresanu D, Moessler H. A 24-week, double-blind, placebo-controlled study of three dosages of Cerebrolysin in patients with mild to moderate Alzheimer's disease. *European Journal of Neurology*. 2006; 13:43–45. [PubMed: 16420392]



155. ClinicalTrials.gov. March 28, 2008  
<http://clinicaltrials.gov/ct2/show/NCT00471211?term=PBT2&rank=1> Viewed
156. ClinicalTrials.gov. March 28, 2008  
<http://clinicaltrials.gov/ct2/show/NCT00117403?term=vitamin+e++CoQ&rank=1> Viewed
157. ClinicalTrials.gov. March 28, 2008  
<http://clinicaltrials.gov/ct2/show/NCT00440050?term=Docosahexaenoic+Acid&rank=2> Viewed
158. ClinicalTrials.gov. March 28, 2008  
<http://clinicaltrials.gov/ct2/show/NCT00088673?term=tramiprosate&rank=2> Viewed

**Table I**

AD Drugs Randomly Selected for Review and, When Available, Publications

Approved Drugs	Drugs Failed or Abandoned in Development	Studies of MCI	Drugs Currently in Active Development
Tacrine [126-128] Donepezil [129-131] Revastigmine [132] Memantine [121,133,134]	Ipidacrine, Sabeluzole [135], ABT-418 [136], Milameline [137], Alvimeline [138], Zifrosilone [139], Vitamin E [117], CX 516 [140], Oxiracetam [141], Propentofylline [118], SC-110, Alfoscerate [120], XE991, Diethylspermine, EGb 761 [119], Neotrofin, Tesofensine, PPI 1019.	Donepezil [142,143] Rivastigmine [108] Galantamine [144,145] Celecoxib [146] Piracetam [147] Rofecoxib [148] Estrogen & Progesterone [149] Aspirin & vitamin E [124]	Huperzine A [150] ACC-001 [151], TTP 488 [152], MCP-7869 [153], MK 0752, Cerebrolysin [154], PBT2 [155], Vitamins E & C, alpha-lipoic acid, & coenzyme Q [156], Docosahexaenoic acid [157], Tramiprosate (Alzhemed) [158].

Legend for Table I: Drug and [references to specific publications included in this study]. When no references are provided this indicates the drug was randomly chosen as an identified AD development candidate but lacks publications reporting outcomes from the intended research.

Table II

**Table II: 1. Drug, Mechanisms, Pharmacological Activities, Design, and Publication<sup>a,b</sup>**

Methodological Query Is there discussion of...? Or relevant data reported about...?	Literature References	Approved AD Studies % Compliant	Failed AD Studies % Compliant	MCI Studies % Compliant	AD Drugs in Development Studies % Compliant
1. Why mechanism is investigated*	[Authors]	90%	39%	70%	50%
2. Drug Class Actions	[Authors]	70%	39%	70%	50%
3. Class Efficacy Evidence	"expect... with previously failed drugs with mechanism, ... specific justifications for why this compound ... a possible exception." [Authors]	90%	39%	70%	50%
4. Why Drug Molecule was Chosen*	"independence of investigators, pharmaceutical funding, publication bias, [6, pp 124-125, 136]	100%	39%	70%	50%
5. Brain Drug Target Concentration Evidence*	Importance of animal studies for dosing to provide appropriate concentrations of drug at brain targets [1,5]	0%	11%	10%	10%
6. Use of Microdosing Studies	Confirm drug concentrations at brain targets [1,5]	0%	0%	0%	0%
7. Other PK evidence considered?	"drug kinetics and dynamics critical to effective use of metrifonate" [1]	0%	6%	10%	10%
8. Elderly PK/PD effects considered?	"changes with aging that potentially affect drug concentrations at molecular targets" [Authors]	0%	6%	10%	10%
9. Design Disease Modifying*/ Symptomatic/Not an Efficacy study (%)	"disease modifying drug effect designs untested" [8-10]	100% 0/100/0	39% 1/1/0	100% 100/0/0%	100% 70/20/0% (One study no information)
10. Statistical analysis ITT specified	"FDA standard" [64]	100%	100%	100%	100%
11. Percent of Studies in Category that are published	"unsuccessful preclinical and clinical developmental studies not published and sponsors refuse to release information about the development." "lack of access to failed studies has unknown effects on understanding how methodologies and practices contribute to drug development failures" [1,5,16-18].	90%	22% /	70%	0% (CTs only registered, plans not published)

**Table II: 2. Dosing<sup>a,b</sup>**

Methodological Query Is there discussion of...? Or relevant data reported about...?	Literature References	Approved AD Studies	Failed AD Studies	MCI Studies	AD Drugs in Development Studies
12. Phase II Evidence	[Authors]	10%	6%	20%	10%
13. Dosing Optimized by studies?	[Authors]	0%	0%	10%	0%
14. Dosing Confirmed in CT?	[Authors]	0%	6%	20%	0%
15. Are Dosing Issues Discussed?*	[Authors]	20%	0%	10%	0%

**Table II: 2. Dosing<sup>a,b</sup>**

Methodological Query Is there discussion of...? Or relevant data reported about...?	Literature References	Approved AD Studies	Failed AD Studies	MCI Studies	AD Drugs in Development Studies
16. Compliance Noted in for this Study?*	[Authors]	20%	11%	20% MEAN = 86/89	0%
17. Basis of Dosing increases during Study	[Authors]	90% 70% by protocol	39% 100% by protocol	100% 100% by protocol	20% 20% by protocol
18. Percentage of Subjects at Target Dose *	[Authors]	60% MEAN = 91%	11%	80% RANGE = 0-70%	10%
19. Interference by Drug Tolerance *	[Authors]	20%	11%	0%	10%
20. Adverse Event Rate Drug/Placebo	[Authors]	80% MEAN = 76/69	6%	40% MEAN = 91/85	0%
21. Discontinuation Rate due to Adverse Events Drug/Placebo	[Authors]	80% MEAN = 18/16	22% MEAN = 11/6	30% MEAN = 15/7	0%

**Table II: 3. Research Subjects<sup>a,b</sup>**

Methodological Query Is there discussion of...? Or relevant data reported about...?	Literature References	Approved AD Studies	Failed AD Studies	MCI Studies	AD Drugs in Development Studies
22. Consideration of Sample Homogeneity & Equivalence across Studies *	"Needed revisions to AD diagnostic criteria" [4]; "population selections, and other factors" [6, pp 124-125, 136]; "diagnostic eligibility vagaries introduce subjects incapable of responding to treatment because of misdiagnosis, genetics, or specific pathology" [25]. "MCI criteria not stable across administrations" [55,115]	100%	39%	100%	22%
23. Presence of other diagnoses *	"whether other diagnoses and medications lead to adverse event reports that suppress dosing increases?" [1]	30%	6%	10%	0%
24. Presence of other medications *	"whether other diagnoses and medications lead to adverse event reports that suppress dosing increases?" [1]	60% MEAN = 88% of subjects	6%	20%	0%
25. Imaging at Entry	"Dubois et al. [4] anticipate imaging becoming part of AD diagnostic criteria"	10%	0%	20%	30%
26. Imaging for Fronto-temporal Dementia (FTD) Exclusion *	"NINCDS Criteria do not exclude FTD" [26]	0%	0%	0%	0%
27. Imaging as Diagnostic Confirmation *	"Dubois et al. [4] anticipate imaging becoming part of AD diagnostic criteria"	0%	0%	0%	0%

**Table II: 4. Outcome Measures<sup>a,b</sup>**

Methodological Query Is there discussion of...? Or relevant data reported about...?	Literature References	Approved AD Studies	Failed AD Studies	MCI Studies	AD Drugs in Development Studies
28. Outcome Measures Specified	[Authors]	100%	33%	100%	50%
29. Rater Qualifications *	"Kobak et al.[33] report that among 29 raters at 12 research sites 72% first learned to administer the Hamilton Depression Scale in	0%	0%	10%	0%

**Table II: 4. Outcome Measures<sup>a,b</sup>**

Methodological Query Is there discussion of...? Or relevant data reported about...?*	Literature References	Approved AD Studies	Failed AD Studies	MCI Studies	AD Drugs in Development Studies
	preparation for the investigation and only 38% were observed using the scale with patients"				
30. Rater Reliability Training *	"problems of inaccuracy, imprecision, and bias interfere with CTs dependent on clinical ratings of outcome measures." [1,15,21,23,25,39,40].	0%	V	20%	0%
31. Ratings Accuracy Verification *	"inaccuracies decrease detected active drug-placebo group differences, undermine statistical significance, and reduce effect size" [15,27].	0%	0%	0%	0%
32. Ratings Precision Verification *	"imprecision increases the variance in data, undermines statistical significance, and reduces effect size" [3,15,21,23].	0%	0%	0%	0%
33. Documentation of Reliabilities Established during Training and Study	"Kobak et al. [33] and Targum [38] showed the effects of more intensive training of investigators at research sites." "More reliable outcome measures can reduce sample size requirements." [30]	0%	0%	10%	0%
34. Rater Bias in Placebo Improvements *	"bias, indicated by placebo group improvement, potentially decreases the detected active drug-placebo group differences and increases the variance in data" [3,15]	0%	0%	0%	0%
35. Are there Rater Disqualifications? *	Demitrack et al. [35] and Kobak et al. [37] found some raters can not qualify in training.	0%	0%	0%	0%
36. Are Means for Multiple Data Points Used for Statistical Analysis? *	"Means of data points reduce variance" [3,21,22]	0%	0%	0%	0%
37. Are Surrogate Outcome Measures (SOM) Used? *	"Rating scale unreliability requires more reliable indicators" [1,4,62]	0%	6%	20%	60%
38. Are SOM Validated? *	"Quantitative SOM are not yet available" [61,61]	0%	0%	0%	10%

**Table II: 5. Research Sites and Investigators<sup>a,b</sup>**

Methodological Query Is there MENTION/discussion of...? Or relevant data reported about...?*	Literature References	Approved AD Studies	Failed AD Studies	MCI Studies	AD Drugs in Development Studies
39. Study number (N) of subjects	"Size issues: [6, pp 124-125, 136]. Aisen & Vellas acknowledged large number of subjects as problems" [12]	100% RANGE 93-632 MEAN = 391	39% RANGE 6-341 MEAN = 150	100%; RANGE 270-6377; MEAN = 1627	100%; RANGE 48-950; MEAN = 292
40. N % lost during study Drug/placebo *	"high dropout rates capable of introducing bias, (JOM 2007, pp 124-125, 136). Long-term disease modifying designs risk losses of subjects" [9,10,12-14]	100% MEAN 27/20	39% MEAN = 14/10	70% MEAN = 27/23	20%;
41. Number of Sites *	"Larger numbers of subjects require greater numbers of clinical research sites." Paul Aisen and Bruno Vellas	80% RANGE = 18-50 MEAN = 32	22% RANGE = 1-28 MEAN = 11	60% RANGE = 6-69 MEAN = 36	90%; RANGE = 1-67 MEAN = 21

**Table II: 5. Research Sites and Investigators<sup>a,b</sup>**

Methodological Query Is there MENTION/discussion of...? Or relevant data reported about....?	Literature References	Approved AD Studies	Failed AD Studies	MCI Studies	AD Drugs in Development Studies
	noted problems with large numbers of subjects requiring "70-80 sites" [12]				
42. No. of Sites < 6 subjects	"Aisen and Velas acknowledged as problems sites contributing only 2-5 patients." [12]. [Authors] "concerned that small sites inexperienced with AD"	0%	0%	0%	10%;
43. Documented site Diagnostic experience *	"Rater bias leads to lax applications of diagnostic criteria to secure subjects" [33]	0%	0%	0%	0%
44. Site Investigator Training Recorded *	"Conditions of research an essential aspect of effective clinical use of drug and evaluations of CT reports." [1]	0%	0%	0%	0%
45. Inter-site Variance Effects on Statistical Analysis Studied *	"inter-site variance undermines CTs power" [13,14]. Acknowledged in Butcher [12] by Aisen & Velas, "large degrees of variance amongst sites." [35]	0%	0%	0%	0%
46. Site Performance in Earlier CTs Stated.	"Proposal to test rater quality" [1]	0%	0% 6% USED SITES EARLIER	0%	0%
47. Site Monitoring Described & Adequacy Evaluated *	"difficulties monitoring sites for conformity to protocols, undermine CTs" [13,14]. "ites falsifying data" [Authors]	0%	0%	10%	0%
48. Intra-study Rater Training provided *	[Authors] "observations in confidential reviews of CTs"	0%	0%	10%	0%
49. Advertising for Subjects Addressed	"Question of effects from advertising compared to self-selected patients" [1]	0%	0%	20%	0%
50. Sites Noted as Commercial or Not	"Question of possible effects from site priorities" [1]	0%	0%	0%	0%
51. Sites Noted as Academic or Not	"Question of possible effects from site priorities" [1]	0%	0%	0%	0%

**Table II: 6. Protocol Controls of Methods and Clinical Transfer<sup>2,4b</sup>**

Methodological Query Is there discussion of...? Or relevant data reported about....?	Literature references	Approved AD Studies	Failed AD Studies	MCI Studies	AD Drugs in Development Studies
52. Are Critical Procedures Controlled by Protocols *	"latest and most rigorous methods for designing and executing study protocols" [6, p. ix]. "lack of adherence of raters and administrators to protocols" [authors]	10%	6%	10%	0%
53. Are Protocols Referenced for SOM *	"Authorities urge caution using surrogate markers to infer clinical outcomes" [99-101]. "Operational protocols needed to overcome human and machine errors" [55,61,62]	0%	6%	0%	10%

**Table II: 6. Protocol Controls of Methods and Clinical Transfer<sup>2,4,b</sup>**

Methodological Query Is there discussion of...? Or relevant data reported about...?	Literature references	Approved AD Studies	Failed AD Studies	MCI Studies	AD Drugs in Development Studies
54. Are Conditions for Effective Drug Use in Clinic Tested	"Knowledge transfer to clinic" [1,102,107]	0%	6%	0%	0%
55. Are Conditions for Effective Drug Use in Clinic Described for Practitioners	"Knowledge transfer to clinic" [1,102,107]	0%	6%	0%	0%
56. Are Clinical Decision Rules Developed and Tested	"clinical decision rules can provide effective guidance for health care policy making and management" [102-107]	0%	0%	0%	0%

<sup>1</sup> As in other categories publications referenced may not be pivotal. For failed AD drugs no publication identified itself or could be identified by the authors as pivotal to the decision to abandon development.

<sup>2</sup> The primary justification for human research in drug development is inform patient care (Kralawish's 2006; Becker and Greig 2008; Kent and Hayward 2007; Calif 2005; Tunis et al. 2003; Rich 2002; Becker 2002); however, transfer of CT conclusions to the clinic seldom addresses the methodological limitations in CTs that also affect patient care (Becker and Greig 2008).

<sup>a</sup>References [Authors] indicate that the subject of the query was developed by the authors without reference to specific publications.

<sup>b</sup>The % Compliant refers to the percent of drug developments compliant with the query aims. The % Compliant does not indicate the percentage of published reports compliant with the aims of the query since failures to publish in the plans for development, or about status or outcomes makes the drug development non-compliant with the aims or standards expressed in the query item.

\* Indicates conditions directly impacting on power.