University of Pennsylvania
ScholarlyCommons

Center for Bioethics Papers

Center for Bioethics

2002

The Search for a Coherent Language: The Science and Politics of Drug Testing and Approval

Jason Karlawish
University of Pennsylvania, jason.karlawish@uphs.upenn.edu

Follow this and additional works at: http://repository.upenn.edu/bioethics_papers

Part of the Bioethics and Medical Ethics Commons, and the History of Science, Technology, and Medicine Commons

Recommended Citation
Karlawish, J. (2002). The Search for a Coherent Language: The Science and Politics of Drug Testing and Approval. Retrieved from http://repository.upenn.edu/bioethics_papers/65

Postprint version. Published in:
Karlawish J.H.T.: The search for a coherent language: The science and politics of drug testing and approval. In: Kapp M.B., ed. Ethics, Law and Aging Review. New York, NY: Springer Publishing Company 8;2002:39-56
The original publication is available at:
http://www.springerpub.com/product/9780826116420

This paper is posted at ScholarlyCommons. http://repository.upenn.edu/bioethics_papers/65
For more information, please contact repository@pobox.upenn.edu.

The Search for a Coherent Language: The Science and Politics of Drug Testing and Approval

Keywords
Alzheimer's disease, treatment, drug development, trial design, medical ethics

Disciplines
Bioethics and Medical Ethics | History of Science, Technology, and Medicine

Comments
Postprint version. Published in:

Karlawish J.H.T.: The search for a coherent language: The science and politics of drug testing and approval. In: Kapp M.B., ed. Ethics, Law and Aging Review. New York, NY: Springer Publishing Company 8;2002:39-56

The original publication is available at:
http://www.springerpub.com/product/9780826116420

This book chapter is available at ScholarlyCommons: http://repository.upenn.edu/bioethics_papers/65

The search for a coherent language: The science and politics of drug testing and approval.

Jason Karlawish, M.D.

Published in:
Ethics, Law and Aging Review. New York, NY: Springer Publishing Company, volume 8; 2002: 39-56.

University of Pennsylvania, Department of Medicine, Division of Geriatrics, Alzheimer's Disease Center, Center for Bioethics, and Leonard Davis Institute for Health Economics.

Address: Institute on Aging, 3615 Chestnut Street, Philadelphia, PA 19104

telephone: 215-898-8997; fax: 215-573-8684;

email: Jason.karlawish@uphs.upenn.edu

"*We can only hope that out of this controversy will come some real advances in the management of a common and dreadful disease*."

Arnold Relman, M.D., Editor, New England Journal of Medicine. January 31, 1991.

The subject of this chapter is the history of the development of treatments for the most common cause of dementia: Alzheimer's Disease. By the middle of the 1980's, industrialized countries recognized that Alzheimer's Disease was common, untreatable and caused substantial harms and costs to individual patients, their family caregivers and society. Moreover, demographic projections of the growth in the elderly population predicted that these harms and costs would grow.[1] Treatments were urgently needed, and the clinical trial was the way to discover them. Years later, clinical trials have yielded results. The 2001 American Academy of Neurology's Evidence based guidelines recommend acetylcholinesterase inhibitors and vitamin E as treatments for Alzheimer's Disease.[2] Progress had occurred.

But the same Academy Guidelines that clearly state how these treatments have value also contains an appendix that acknowledges there is no standard approach to determining the magnitude of the benefits of dementia treatments.[2] Other publications report this and other limitations of Alzheimer's Disease treatment trials. Common themes are disagreements over the choice of endpoints, and the interpretation of treatment versus control group differences in endpoints.[3-11] The message is that the communities of clinical medicine, clinical research, research regulation and industry have not achieved a set of measures or approaches to analyze them that fulfill both the

standards of valid science and clinical value. The result is an incoherence in the language of benefit. The term "incoherence" describes the inability of these communities to select and analyze measures of treatment benefit so that claims of validity and clinical value hang together.

A language of benefit serves a number of interrelated ethical and scientific functions. Hence, the costs of the failure to achieve a coherent language are significant. Clinical investigators and institutional review boards struggle with weighing research risks and benefits. Clinicians and patients struggle with the decision whether to use a new treatment and the merits of changing or discontinuing treatments. Finally, public and private policy makers cannot reach transparent and acceptable decisions about access to and reimbursement for treatment. There is also a human cost. As long as incoherence exists, clinical research results are of dubious value. As a result, further clinical trials are done to try and achieve coherence. But these trials require significant commitments of financial, scientific, regulatory and human resources. Of particular concern is the cost of human resource: namely, human subjects, especially elderly persons with chronic and ultimately fatal illnesses.

The purpose of this chapter is to examine the interrelated scientific and ethical issues that are the origins of how languages of treatment develop. The public image of treatment development is that it follows a "rational path" that begins with preclinical investigations and progresses logically with systematic and careful studies in humans (see for e.g.).[12] These clinical trials use the most valuable of valid measures to establish efficacy. Experts restrain their enthusiasm for encouraging reports until results are replicated [13]. Experts are not influenced by an "emotional side to the story."[14(pg. 528-529)] In

this model, validity and value are distinct and separate concepts that are adjudicated by a disinterested community of scientists.[15] A study is either valid or it is not, and validity cannot be "traded off" for value.[16]

But the history of Alzheimer's Disease treatment development suggests otherwise. The approach to analyzing this history relies on analyzing the actual, historically situated social interactions and causal routes that were involved in the development of Alzheimer's Disease treatments.[17(pg. 48)] The lessons are that Alzheimer's Disease treatments were not inevitable and need not be the way they are. They are the result of a matrix of intersecting interests and ideas. Lessons of this matrix include that ideas about what is Alzheimer's Disease interact with the choice of measures of Alzheimer's Disease treatment efficacy, and that disagreements about validity and value tend to overlap into one and the other and tradeoffs occur between them. Resolutions of these disagreements largely follow the lines of authority and the power attached to them. What follows is the history of Alzheimer's Disease treatment development. This history will focus on the development of symptomatic and disease slowing treatments for Alzheimer's Disease.

The history of Alzheimer's Disease treatment.

Alzheimer's disease is a chronic, progressive and ultimately fatal neurodegenerative dementia. Typical symptoms include impairments in short-term memory, language, personality, and the abilities to organize tasks and spatial arrangements. Diagnostic criteria require deficits in two or more distinct cognitive functions, such as verbal memory and executive function, that are significant enough to

impair a person's ability to perform their usual and everyday tasks such as managing finances or cooking.[18] In addition, because the disease often impairs a patient's insight and judgment into the scope and severity of their symptoms, clinicians typically obtain collateral history from a knowledgeable informant. The key point here is that the diagnosis of Alzheimer's Disease relies on assessing the changes in each patient's baseline function and then linking these changes to declines in at least two cognitive functions. The criteria do not describe a uniform set of measures of function and cognition to diagnose Alzheimer's Disease or the relative weights to apply to history obtained from the patient versus the knowledgeable informant.

The staging of the disease is similarly multi-factorial in both what is measured and the sources of information. Criteria specify assessment of a number of broadly defined domains such as "judgement and problem solving" and either require a global judgment of severity or use a weighted algorithm.[19] Notably absent from staging criteria are behavioral disorders such as agitation and psychosis because over the course of the disease these symptoms wax and wane.

In summary, while the criteria for both the diagnosis and staging of the disease show good reliability and validity, they require clinicians to assess a number of cognitive domains, make judgments about the degree of impairment in them and then weight these judgments into a final assessment. No one measure or set of measures defines both that a person has Alzheimer's Disease and how severe it is. In addition, patients, caregivers and clinicians differ in the kinds of symptoms that matter to them.[20, 21] As a result, when investigators began to design clinical trials to test whether an intervention treats the symptoms of Alzheimer's Disease, they found that standards to diagnose and stage the

disease did not readily translate into measures of treatment benefit. A new language was needed.

Discovering a language of symptomatic benefit -- the acteylcholinesterase inhibitors: The acetylcholinesterase inhibitors inaugurated the clinical science of Alzheimer's Disease drug development. As of 2001, four FDA approved medications are marketed as safe and effective treatments for the symptoms of mild to moderate Alzheimer's Disease. In order of their approval, they are tacrine, donepezil, rivastigmine, and galantamine.

Development of these drugs began in earnest in 1986, after the New England Journal of Medicine published the results of Summers' and colleagues clinical trial that measured the symptomatic benefit of tacrine in 17 patients with Alzheimer's Disease.[22] The design involved a series of within subject controlled and open-label phases and measured efficacy using both cognitive measures and "the daily global assessment." The authors concluded: "The degree of improvement has often been dramatic. One subject was able to resume most of her homemaking tasks, one was able to resume employment on a part-time basis, and one retired subject was able to resume playing golf daily."[22 (pg. 1243)] Although the drug required four times a day dosing, caused annoying side effect such as nausea and diarrhea and required regular monitoring of liver enzyme activity, these reports of dramatic functional improvements on the global measure suggested the risks were worth the benefits. An accompanying editorial described the study as a step along a "rational path" and praised it as "a triumph for the scientific method."[12(pg. 1287)]

The research inspired calls for rapid follow up studies to confirm the drug's effectiveness as a treatment for Alzheimer's Disease.

Five years later, the *New England Journal of Medicine* revisited the Summers' study with a "Special Report." This series of three unusual articles included the Journal's editor Arnold Relman describing the controversy that surrounded both the decision to publish the original article and the FDA's subsequent investigation of Summers' study, the FDA's summary of the violations by Summers, and a response by Summers and colleagues.[23] The critical issues FDA identified included basic failures of design and conduct such as randomization, blinding and accurate recording of the global measure. The FDA concluded the evidence to be the equivalent of uncontrolled and anecdotal information.[13]

Relman defended the decision to publish the study. He disclosed that both reviewers and the editors found the study imperfect and preliminary using methods not as rigorous as might be desired. Relman quoted a referee to defend the decision to publish: "The author's results should encourage further studies, which in itself is reason for publication." Reviewing this decision, Relman reflected, "We can only hope that out of this controversy will come some real advances in the management of a common and dreadful disease."[23(pg.349)]

The Summers study has at least two main lessons. First, it illustrates how considerations of scientists' perception of public desperation and hope as well as their own experience of these emotions alter the threshold of scientific validity. Summers, the editors at the *New England Journal of Medicine*, peer reviewers, and editorialists were all willing to relax standards of clinical trial review because they felt that Alzheimer's

Disease is a common and dreadful disease and the scientific and clinical communities needed encouragement. But alternatives existed. For example, the Journal could have rejected the manuscript. This study could have joined previous studies of tacrine by Summers that were published in less prestigious journals.[24] Or, the editors could have commissioned or themselves written an editorial that candidly presented the controversy surrounding the publication decision and admitted the motivation to inspire future study. This illustrates the second part of this lesson: the scientific community is unwilling to publicly admit that these considerations influence their decisions. Instead, the editorial concluded the study fit as the next logical step in a rational path.[12] Were it not for the FDA investigation and a settlement agreement that included publication of the investigation and Summers' response in a medical journal, the controversy would never have been known to the public or to the scientific community.

The second lesson of the Summers study is in its measures of efficacy. In this trial, the measures of efficacy were a combination of cognitive tests (names learning and orientation tests) and a measure called "the daily global assessment." The citation for the global assessment was a previous study of tacrine by Summers in 12 patients with Alzheimer's Disease that simply states that "additional monitoring of THA effects were done by a global physician assessment and evaluation of nursing notes." [24] Despite this ambiguity, this global measure provided the data for the vivid narratives of clinical response that inspired the follow up studies. But the measure was unknown. The subsequent studies of tacrine did not use Summers' measures, but the study inaugurated the debate over what measures of clinical response should be.

The effect of the Summers' trial was a rapid sequence of events that led to an effort that combined the financial and patient resources of the Alzheimer's Association, National Institute on Aging, Warner-Lambert, and the Food and Drug Administration (FDA) to design and conduct clinical trials of tacrine. The three multi-site, placebo controlled randomized trials of tacrine followed the controversial 1986 *New England Journal of Medicine* study illustrated the challenge of measuring a language that unified both valid and valuable efficacy data for the treatment of Alzheimer's Disease.

In the first study, subjects who received tacrine compared to those who received placebo showed a statistically significant 2.4 point difference in the mean performance on the measure of cognition, but unlike Summers' promising study, did not show significant differences on the global measure of change.[25] An accompanying editorial pronounced the end of the cholinergic hypothesis for symptomatic benefit [26] and reinforced this view in response to critical letters to the editor.[27] In an explicit recognition of the resource issues at stake, the editorialist wrote "I believe that time, effort, and money will be better spent in developing treatments that attack fundamental problems of neuronal degeneration than in designing a 'better' acetylcholinesterase inhibitor."[27(pg. 810)]

But the nascent Alzheimer's Disease research community declined this shift in resources. In both the article's "discussion" and the hearings of the FDA's Peripheral and Central Nervous System Drugs Advisory Committee that reviewed the results, the debate centered on the interpretation of the mixed results. The article described both the measure of cognition and the global as the primary measures, but its discussion section suggested that the changes on the cognitive measure were sufficient to establish that tacrine was effective. Representatives of Warner-Lambert were more forthright in this

interpretation. They argued that the study was built around the ADAS and its cognitive subcomponents and this outcome measure deserves the closest attention.[14(pg71)] Ultimately, the Advisory Committee judged the mixed result inconclusive and voted not to recommend tacrine for approval. Additional data reviewed four months later failed to change the decision.[28]

The search for the perfect language -- the global measure: The tacrine trials raised a number of issues specific to tacrine (including the proper doses of drug, the use of cross-over design, the duration of washout periods). They also raised the interrelated scientific and ethical issues that are the origins of creating a coherent language of benefit: the choice and interpretation of measures of a treatment's benefit.

The failure of the first pivotal trial was a disappointment and also a surprise. Summers reported that a global measure detected dramatic and frankly poignant responses to tacrine: retirees playing golf and resuming part-time work. In contrast, the subjects in the subsequent trial failed to show responses on the global measure. General consensus was that the measure of cognition (the Alzheimer's Disease Assessment Scale or ADAS) was a valid measure of changes in patient cognitive function. But the ADAS was a long scale that was not used routinely either by clinicians or even most Alzheimer's Disease Centers. The statistically significant differences between the average scores of control and intervention groups were difficult to translate into a valued measure of symptomatic benefit. In short, did clinicians notice a 2.4 point change? Statistically significant differences were also found in measures of functional decline, but these changes raised the same question: were they clinically significant? Attention turned

to the choice and interpretation of measures to establish "clinically valuable change." Specifically, the debate turned on two issues: the structure of the global assessment, and the role of the caregiver in the assessment of change.

The basic structure of a global measure of change is that a rater assesses a patient at baseline and then after some time interval reassesses the patient.[29, 30] The rater compares baseline and reassessment data to judge whether the person has stayed the same or has had a clinically significant "marked," "moderate," or "minimal" decline or improvement. There was general agreement that these qualities substantiate why the measure is "real world," "ecologically valid," and "holistic."[29-31] But there was significant disagreement about how to perform a global measure.[29, 30] A review of these disagreements shows confusion in the distinction between what is a problem of validity versus what is a problem of value. Ultimately, while science was the playing field for these disagreements, the resolution followed lines of authority and power.

A general principle of instrument assessment is that reliability precedes validity.[32] That is, an instrument must be reliable in order to be valid. In the case of the global, evidence suggested that the measure of global change had only fair reliability.[33] This largely reflected the lack of structure to the measure. It did not specify the domains a clinician should assess, how to rate them or how to weigh them. Basic principles of measurement science are that increasing structure improves a measure's reliability. But greater structure has another effect that runs contrary to the value of the global: it would diminish the measure's face and content validity. Specifically, structure increases sensitivity to detect change, but this change might be clinically insignificant and not

reflect clinically significant change: the property was at the heart of the global's value as a measure of treatment efficacy.[29, 30]

A second controversy was the sources of information for making a global rating.[29, 30] The focus was on the role of the patient's knowledgeable informant, typically a family caregiver. There was general agreement that the baseline interview should include interviews with patient and caregiver, but there was disagreement about whether the rater should interview the caregiver in the follow-up interviews.

FDA offered two arguments against caregiver input. First, its statutory mandate required that measures of treatment of the disease should be separate from measures of safety. Caregiver input would introduce information about side effects.[29] These data would confound an assessment of efficacy. In other words, safety data (such as stomach upset) would contaminate an assessment of improvements in symptoms. Second, officials argued on the basis of effect sizes. A skilled clinician who alone detected change using an unstructured global would detect significantly large enough change to say that a drug has clinically meaningful benefit. An FDA official explained:

> Those of us in the FDA who were interested in an effect-size issue – which I don't think we've solved, and that speaks to clinical significance – wanted very much to have something that would allow us to detect, independently of somebody's arbitrary set of rules, something that was large enough for clinicians to see; the argument being, if they could see it, at least that was a minimum standard. [30]

In response, clinical investigators argued that the practice of Alzheimer's Disease clinical care relied on an interview with knowledgeable informants, especially the patient's caregiver. This interview disclosed important information such as patient behavior and the abilities to perform everyday activities of daily living. Absent these information, the clinician's global would simply be a global assessment of patient cognitive function.[21] To deprive a clinician access to this information during a global rating hindered the ratings face and content validity. These clinicians maintained that the clinician rater could be kept blinded to reports of side effects and forbidden from inquiring about them during follow-up interviews.

The debates over the degree of structure and role of the caregiver show a tension between trying to simultaneously maximize a measure's validity and value. For example, the argument for access to the caregiver featured improving the measure's face and content validity, which in turn improved its value. In contrast, the argument against access focused on the impact on the value of measuring small changes. The argument for more structure appealed to improving the measure's reliability. In contrast, the argument against more structure focused on the limited value of detecting trivial effects. In sum, the measure could not be maximally valid and valuable, but only valid enough for a given sense of value.

At the core was a disagreement about what kinds of symptoms reflect Alzheimer's Disease and thus the best measure to document improvement in the symptoms of Alzheimer's Disease. One view of Alzheimer's Disease is that measures of cognition are the best expression of an "antidementia effect" because Alzheimer's Disease is a cognitive disorder.[9, 34] Hence, improvements in measures of cognition reflect

improvements in the disease. Thus, a clinician's global without caregiver input is a valid and valuable measure of change. Another view of Alzheimer's Disease is that measures of functional impairment are the best expression of Alzheimer's Disease.[35] Hence, improvements in measures of function reflect improvements in cognition. Thus, a global measure should include caregiver input or even be exclusively made by the caregiver. In short, ideas about what is Alzheimer's Disease impact upon judgments about what is the best measure of Alzheimer's Disease treatment.

The outcome of these debates was that subsequent trials of acetylcholinesterase inhibitors used a semi-structured global with caregiver input. The interviewer was instructed not to inquire about side effects. Some trials included a caregiver rated global, but this was as a secondary endpoint.[36] This outcome largely followed along the lines of power and authority. FDA championed the unstructured interview that excluded the caregiver. [29] Officials advocated for this global on the basis of effect sizes. But this position exceeded their authority. Approval focuses on whether statistically significant differences exist between treatment and control, not on the size of the difference. The expert medical community defines what is a clinically significant effect and how to measure it, and the expert community wanted a semi-structured measure that included an interview with the caregiver.[37] But did this measure translate into a coherent language of treatment? The deliberations of the expert community suggest the answer to this question was a qualified "maybe."

Translating the language of symptomatic benefit – the approval of tacrine. The controversies over the measures of benefit used on the three pivotal tacrine studies were

on full display at the third meeting of the Peripheral and Central Nervous System Drugs Advisory Committee. The committee's decision would not determine the approval tacrine. Only FDA had that authority. But its decision would strongly influence the FDA's decision and thereby set the standard for the design and review of future symptomatic treatments for dementia. The FDA's instruction to the committee was to answer the "critical question" whether substantial evidence supported that tacrine was effective. FDA emphasized that this was not a statistical issue but a clinical judgment by expert professionals.[38] It was, in effect, a judgment of the value of the measured benefits of tacrine. The committee members and invited guests drawn from the leaders of neurology clinical research and Alzheimer's Disease clinical trials (including investigators from previous tacrine trials) were being asked to view the evidence from a different perspective: the perspective of clinicians and patients.

The committee members' discussions showed general agreement that tacrine had a statistically significant effect on the primary endpoints. But the members took sides on how to make the judgement that these effects constituted a treatment benefit. In a phrase that recurred throughout the meeting, the members struggled to translate these data into "real world" language. Two issues articulated this struggle and the positions of the different sides: the value of the primary endpoints, and how to interpret the results of the global measure. Despite these issues, the committee voted unanimously to recommend approval and FDA ultimately approved of tacrine as a symptomatic treatment for patients with mild to moderate Alzheimer's Disease. Several committee members cited Leonard Berg's summary on the evidence as an acceptable resolution of the disagreements. Berg engaged in an imaginary clinical encounter with a patient and family:

> You can't expect it [tacrine] to do very much for you. There is a very small chance that it will help you a good deal, there is a little better chance that there will be some measurable improvement that some people might say is clinically significant. How important that degree of improvement is depends on your perception, not on my perception, and we know that there are many people who will be delighted and call very important something that many of us around the table would consider of little or no impact.[38(pg. 219)]

Since the publication of the Summers study in 1986, seven years of debates about the choice of the endpoints and how to interpret them were effectively set aside: benefit depends on your personal perspective. Berg's summary was as "value neutral" as the committee could be in its efforts to answer a value-laden question. The message was that the committee is in no position to impose a standard of benefit. But, his comment suggested the lingering skepticism over the value of the clinician rated global ("How important that degree of improvement is depends on your perception, not my perception").

Berg offered a follow-up comment that illustrated a problematic consequence of this perspective: "I would also say, because as a treating physician, I have to pay attention to some to cost-benefit ratios, that the cost is substantial not only for family, but for third-party payers and society at large." This point raised the issue of whether the drug was cost-effective. However, FDA approval does not include pharmacoeconomic issues: "But across this consultation table we are talking about benefit versus risks, and

the benefits appear to be small, but the chance of very serious side effect to the drug appears to be small."[38(pp. 219-20)] Lines of authority rendered pharmacoeconomic issues moot, but the comment heralded disagreements that followed the approval of tacrine and other acetylcholinesterase inhibitors.

In countries such as Great Britain and Canada where approval or state reimbursement is linked to assessments of both efficacy and cost-effectiveness, tacrine was not approved and subsequent acetylcholinesterase inhibitors such as donepezil were the cause of considerable debate. For example, in Canada some provinces did not reimburse donepezil while others imposed criteria based on patient baseline mini mental state exam (MMSE) scores and changes in these scores. The irony of these criteria is that the MMSE is one of a few measures that are widely used in Alzheimer's Disease clinical practice, but it only measures cognitive function. Much like changes in the ADAS-cog, the clinical significance of a few points of change is opaque. After seven years of research, the language of benefit in clinical trials to develop Alzheimer's Disease symptomatic treatments was adequate for a panel of experts but it remained difficult to translate into the language of clinical practice or policy.

Discovering a language of disease progression -- the vitamin E debate: Acetylcholinesterase inhibitors were developed according to a model of symptomatic benefit. But the accumulated evidence of the many trials generated a discussion: do these drugs slow the progression of disease? The FDA Advisory Panel that voted to approve tacrine dismissed it as a topic for review. A member remarked "I would submit there is

no way to give an answer to that at the present time and we shouldn't waste any time discussing it."[38(pg. 219)] But the issue did not disappear.

One consequence of the acetylcholinesterase studies was that they assembled an infrastructure of university based Alzheimer's Disease clinical investigators that organized into the NIA-funded Alzheimer's Disease Cooperative Study (ADCS) whose mission included studying potential treatments for Alzheimer's Disease, especially treatments that did not have the support of industry. The ADCS' first clinical trial was a 24 month long placebo controlled study of the individual and combined effects of vitamin E and selegiline on patients with moderate Alzheimer's Disease.[35] The trial's outcome measure was time to the onset of any one of the following events: death, nursing home placement, dependence in two or more of three basic activities of daily living (eating, grooming and using the toilet), and progression to severe stage disease. Efficacy was defined as a significant delay in the time to achieving this combined endpoint. In contrast to studies of acetylcholinesterase inhibitors, measures of cognition were secondary endpoints.

The results published in the April 1997 *New England Journal of Medicine* showed subjects who received vitamin E or selegiline had significantly greater delays in the time before they reached one of the four endpoints and no significant differences in their performance on the secondary measures of cognition. Based on the criteria of primary endpoints, the trial was a success. The authors concluded vitamin E and selegiline slow the progression of Alzheimer's Disease. The media reported the arrival of two new treatments for Alzheimer's Disease.

But the collective results of primary and secondary endpoints presented mixed results. Treatment to control group differences existed in a series of functional endpoints but no differences existed in cognitive measures. The authors' discussion of these results engaged the issue of what is the best measure of disease progression. In a clear recognition and then dismissal of the paradigm developed for acetylcholinesterase inhibitors, they wrote: "Although cognitive measures have typically been the index of symptomatic improvement measured over a short interval, they may not be the best measures of disease progression."[35(pg 1221)] The authors raised the provocative, even revolutionary, point that a "cognitive measure" such as the ADAS may not be the best way to measure effects on the disease. They cited the result of significant treatment to control differences in performance on measures of instrumental activities of daily living that require cognitive function and concluded "Perhaps functional and occupational measures of cognitive capacity are better indicators of disease progression than psychometric measures."[35(pg. 1221)] The authors concluded "In patients with moderately severe impairment from Alzheimer's Disease, treatment with selegiline or alph-tocopheral slows the progression of disease."[35(pg. 1216)] This claim makes two points: the drugs treat Alzheimer's Disease and the measures that established this are measures of disease progression.

Unlike the acetylecholinesterase inhibitor trials, the ADCS study's endpoints were highly coherent across clinical trial and clinical practice. The investigators viewed them as valid and arguably most caregivers and clinicians would accept them as valuable too, although the benefit of survival, of living longer with Alzheimer's Disease, is highly contingent on assessments of patient quality of life.[39] But to the clinical trial world the

study was incoherent. A critical accompanying editorial, "Treatment of Alzheimer's Disease – Searching for a breakthrough, settling for less," questioned the validity of the results. The criticisms included the appropriateness of the choice of the endpoints.[9] While valuable to clinicians, the authors argued they could not accurately measure disease progression.

Their criticism went as follows. Death is not a direct consequence of Alzheimer's Disease. Nursing home placement is more closely related to behavior problems and "the diminished tolerance of the caregivers" than to impaired cognitive ability. The failure to affect cognitive test scores was "perplexing." The editorialists speculated that "Perhaps the treatment effect was only symptomatic, affecting behavior but not the underlying disease process."[9(pg. 1247)] The editorial summarized the criticism: "The composite end point is also an uncertain surrogate for the progression of Alzheimer's disease because the end point may be reached for reasons unrelated to disease progression – for example, an early death due to cancer, or placement in a nursing home because of the illness of a spouse."[9(pg. 1246)]

The editorial was significant not only for what it said but who said it: David Drachman and Paul Leber. Drachman was a co-investigator on acetylcholinesterase trials,[25] co-author of the diagnostic criteria for Alzheimer's Disease,[18] and member of the FDA Advisory Panel that recommended approval of tacrine.[38] Leber was the director of the FDA's neuropsychopharmacology division. He was intimately involved in the history of Alzheimer's Disease treatment development. In effect, a member of the dementia expert community and the head of the FDA were on record saying that the ADCS study was invalid and thus of little or no value.

The political magnitude of the conflict is significant. The ADCS vitamin E trial was not the effort of a small group of investigators and subjects. Twenty-three Federally funded sites at academic medical centers enrolled 341 subjects. The study's authors and co-investigators included many who persons were participated in the acetylcholinesterase trials. Leber, the co-author of the critical editorial, presided over the FDA's development of guidelines for the review and approval of antidementia drugs. In short, the ADCS vitamin E trial and its accompanying editorial displayed the radical difference in the medical scientific community over what measures serve as both valid and valued language of Alzheimer's Disease treatment. Subsequent evidence-based reviews of the study reflect this incoherence.[2, 11, 40, 41]

The current state of Alzheimer's Disease treatment development.

The acetylcholinesterase and vitamin E trials established a worldwide network of Alzheimer's Disease clinical trial centers, industry funding, regulatory standards for review and approval, and a basic template of trial design and analysis. But the products of this effort have been controversial. Three acetylcholinesterase inhibitors are marketed as symptomatic treatment for Alzheimer's Disease, but skepticism lingers over claims of their effectiveness.[3-10] Tacrine has essentially faded from use though its champion remains insistent it is beneficial. As recently as 2000, in the libertarian *Medical Sentinel* (a journal "dedicated to the pursuit of liberty, free markets and integrity in medical research"), Summers summarized the tacrine trials and concluded that more patents should be treated with it.[42]

The "dual endpoint" strategy of a cognitive measure plus a clinician rated global achieves regulatory requirements, but it is not the widely accepted paradigm for trials. European drug regulators add a third endpoint: a measure of function,[43] and trials often have functional measures as primary endpoints and cognitive measures as secondary endpoints.[44-46] The rhetoric of Alzheimer's Disease treatment has transformed from symptomatic response to the preservation of function, that is, patients may not noticeably improve on the acetylcholinesterase, but they will decline slower.[45] Experts hint that the drugs may slow progression (Winblad, Brodaty et al. 2001), though none of the drugs have achieved regulatory approval as disease slowing and no clear consensus exists on how to establish that a treatment slows Alzheimer's Disease.[34]

In sum, the histories of the study of acetylcholinesterase inhibitors and vitamin E have merged. The 1992 message of the critical editorial that accompanied the negative tacrine study that followed up Summers' inspirational but invalid study lingers: "time, effort, and money will be better spent in developing treatments that attack fundamental problems of neuronal degeneration than in designing a 'better' acetylcholinesterase inhibitor."[27(pg. 810)] That is, the focus of research resources ought to be on establishing valid and valuable endpoints to establish disease slowing.

Conclusion.

The clinical trial is widely recognized as the best way to establish the safety and efficacy of treatments. It is the cornerstone for the development of rational therapeutics and the combined forces of evidence based medicine and both public and private managed care regard it as a critical standard to determine what is the standard of care.[47]

This essay has focused on the history of Alzheimer's Disease researchers' efforts to use the clinical trial. While the image of drug development is that of linear progress that begins at the benchside and then moves through progressive phases of clinical research that begin with safety and dose finding studies, the history of Alzheimer's Disease treatment development describes a different process. Four features describe this process.

The first feature is the emotions of the scientific establishment. Researchers and reviewers publicly discuss their role using rhetoric that is measured and dispassionate while subjects are described as desperate, even irrational, in their pursuit of treatment benefit in research. The clinical trial is seen as an essential institution to prevent the rise of irrational therapeutics. But the history of Alzheimer's Disease drug development suggests that researchers and subjects may share more emotions than is typically portrayed. The controversy surrounding the publication of the Summers' study in the *New England Journal of Medicine* is a rare glimpse into the influence of desperation and hope on peer reviewers, editorialists and editors.

The second feature of this process is the breakdown of the validity and value distinction. Validity is portrayed as a threshold concept. A study is either valid or it is not and validity cannot be "traded off" for value. But the history of Alzheimer's Disease treatment development suggests the distinction blurs. Disagreements about validity and value tend to overlap into one and the other and tradeoffs do occur. The history of the global measure and the vitamin E trial suggests that scientists begin with a vision of value and search for a valid measure that supports this vision.

The third feature is the qualities of a coherent language of treatment benefit. Fields such as hypertension, diabetes and AIDS have made rapid progress in the

development of therapeutics because common measures describe both what is the disease and whether it is being treated. For example, in the case diabetes, a clinical trial will measure the subjects' glycosylated hemoglobin A1C concentration (HGA1C). This measure has a number of meanings. It defines diabetes. That is, when it is above a threshold value, a person has diabetes. It also defines the treatment of diabetes. Clinicians and patients follow the value to inform them both of the severity of disease and the success of treatment. The lower the HGA1C, the more successful the treatment. In sum, a HGA1C is both a valid and valued measure. The development of treatments for other chronic diseases such as hypertension and osteoporosis follow a similar model. Blood pressure elevation and reduction of bone mineral density define the presence of disease and changes in these measures define successful treatment. The history of AIDS treatment development shows how these measures can change. Initially, a change in CD4 lymphocyte count defined treatment benefit. Now, changes in the levels of viral load are included in the assessment of treatment benefit.

The point is that ideas about what a disease is interact with definitions and measures of what are effective treatments of that disease. The language to express these ideas works when (1) validity and value are expressed using common terms, (2) researchers, clinicians and patients accept the terms, and (3) clinicians and patients can measure and talk about the language. In the case of Alzheimer's Disease, disagreements about whether Alzheimer's Disease is a "cognitive disorder" or a "functional disorder" led to different ways to talk about what is Alzheimer's Disease. As a result, a language of treatment is difficult to cohere.

The fourth feature of drug development is that it confirms what other histories have shown,[48] that the claim of treatment benefit depends on the network of interested parties that attach the label of "benefit" and that these parties exercise power and persuasion to resolve these disagreements. Three core constituencies are involved in this network: funders (both public agencies such as NIA and private industry such as Warner-Lambert), researchers, research regulators (the FDA). In the case of some disease, such as AIDS, a fourth constituency is involved: patients with the disease.[49] An essential issue negotiated among these constituencies is a coherent language of benefit. The history of Alzheimer's Disease drug development shows how among the constituencies, FDA is a kind of partially shackled giant. The agency has significant control over setting standards for trial design and review, but these standards are constrained by its statutory authority to regulate the marketing claims on industry and responding to the consensus of the "expert community." While FDA certainly influences issues of clinical significance and pharmacoeconomics, these matters are largely left to the community of researchers in collaboration with industry, who, as the first point indicated, are themselves motivated by desperation and hope.

The overall lesson of the history of Alzheimer's Disease treatment discovery is that treatments are not inevitable and need not be the way they are. Thousands of patients have participated in Alzheimer's Disease clinical trials and yet results remain in dispute. Disagreements about the interpretation of research results are framed in the dispassionate language of science but are largely resolved along lines of authority and power. The critic must ask: how well does the system represent the interests of the people it intends to serve?

References

1. Brookmeyer R, Gray S, Kawas C. Projections of Alzheimer's disease in the United States and the impact of delaying disease onset. Am J Public Health 1998;88:1337-42.

2. Doody RS, Stevens JC, Beck C, et al. Practice parameter: management of dementia (an evidence-based review). Report of the Quality Standards Subcommittee of the American Academy of Neurology. Neurology 2001;56:1154-66.

3. Melzer D. Personal paper: New drug treatment for Alzheimer's disease: lessons for healthcare policy. Br Med J 1998;316:762-4.

4. Van Gool WA. Efficacy of donepezil in Alzheimer's disease: fact or artifact? Neurology 1999;52:218-9.

5. Bentham P, Gray R, Sellwood E, Raftery J. Effectiveness of rivastigmine in Alzheimer's disease. Improvements in functional ability remain unestablished [letter; comment] [see comments]. Br Med J 1999;319:640-1.

6. Winker MA. Tacrine for Alzheimer's Disease Which Patient, What dose? Journal of American Medical Association 1994;271:1023-4.

7. Selai CE, Trimble MR, Rossor MN, Harvey RJ. Effectiveness of rivastigmine in Alzheimer's disease. Patients' view on quality of life should be assessed [letter; comment]. Br Med J 1999;319:641-2.

8. Pryse-Phillips W. Do we have drugs for dementia? No. Arch Neurol 1999;56:735-7.

9. Drachman DA, Leber P. Treatment of Alzheimer's disease -- searching for a breakthrough, settling for less (editorial). N Engl J Med 1997;336:1245-7.

10. Post SG, Whitehouse PJ. Emerging Antidementia Drugs: A Preliminary Ethical View. J Am Geriatr Soc 1998;46:784-7.

11. Canadian Coordinating Office for Health Technology Assessment. Technology Report. Drug Treatments for Alzheimer's Disease. II. A Review of Outcome Measures in Clinical Trials; 2000.

12. Davis KL, Mohs RC. Cholinergic drugs in Alzheimer's Disease. N Engl J Med 1986;315:1286-7.

13. Division of Neuropharmacological Drug Products. Office of New Drug Evaluation (I). Center for Drug Evaluation and Review. An Interim Report from the FDA. N Engl J Med 1991;324:349-52.

14. Food and Drug Administration. Peripheral and Central Nervous System Drugs Advisory Committee. Meeting #25. In; March 15, 1991; Rockville, MD; March 15, 1991. p. 1-528.

15. Fried C. Medical Experimentation: Personal Integrity and Social Policy. Amsterdam: North Holland Publishing Company; 1974.

16. Freedman B. Scientific value and validity as ethical requirements for research: A proposed explication. IRB: A Review of Human Subjects Research 1987;9:7-10.

17. Hacking I. The Social Construction of What? Cambridge, MA: Harvard University Press; 1999.

18. McKhann G, Drachman D, Folstein M, Katzman R, Price D, Stadlan E. Clinical diagnosis of Alzheimer's disease. Report of the NINCDS-ADRDA work group under auspices of the Department of Health and Human Services Task Force on Alzheimer's disease. Neurology 1984;34:939-44.

19. Berg L. Clinical dementia rating (CDR). Psychopharmacol Bull 1988;24:637-9.

20. Post SG, Stuckey JC, Whitehouse PJ, et al. A focus group on cognition-enhancing medications in Alzheimer Disease: Disparities between professionals and consumers. Alzheimer Dis Assoc Disord 2001;15:80-8.

21. Claus JJ, Teunisse S, Walstra GJM, Gool WAv. Determinants of global clinical change assessment in patients with early Alzheimer's disease. Dement Geriatr Cogn Disord 1998;9:157-63.

22. Summers WK, Majovski LV, Marsh GM, Tachiki K, Kling A. Oral Tetrahydroaminoacridine in Long-term Treatment of senile demential, Alzheimer type. The New England Journal of Medicine 1986;315:1241-5.

23. Relman AS. Special Report: Tacrine as a treatment for Alzheimer's Dementia. The New England Journal of Medicine 1991;324:349.

24. Summers WK, Viesselman JO, Marsh GM, Candelora K. Use of THA in Treatment of Alzheimer-like Dementia: Pilot Study in Twelve Patients. Biol Psychiatry 1981;16:145-53.

25. Davis KL, Thal LJ, Gamzu ER, et al. A Double-Blind, Placebo-Controlled Multicenter study of Tacrine for Alzheimer's Disease. The New England Journal of Medicine 1992;327:1253-9.

26. Growdon JH. Treatment for Alzheimer's Disease? The New England Journal of Medicine 1992;327:1306-8.

27. Growdon JH. Tacrine in Alzheimer's Disease (letter). N Engl J Med 1993;328:810.

28. Food and Drug Administration. Peripheral and Central Nervous System Drugs Advisory Committee. Meeting #27. In; July 15, 1991; Bethesda, MD; July 15, 1991. p. 1-267.

29. Leber P. Guidelines for the Clinical Evaluation of Antidementia Drugs. 1st Draft. Rockville (MD): Food and Drug Administration, Center for Drug Evaluation and Research, Division of Neuropharmacological Drug Products; 1990 Nov.

30. Schneider LS, Olin JT. Clinical Global Impression of Change. Clinical Global Impressions in Alzheimer's Clinical Trials. Int Psychogeriatr 1996;8:277-90.

31. Doody RS. Test scores in clinical trials vs performance in real life: Can clinical global assessments bridge the gap? In: Wimo A, Jonsson B, Karlsson G, Winblad B, eds. Health Economics of Dementia: John Wiley and Sons Ltd.; 1998:311-25.

32. Streiner DL, Norman GR. Validity. In: Health Measurement Scales: A Practical Guide to their Development and Use. Oxford: Oxford University Press; 1996:144-62.

33. Dahlke F, Lohaus A, Gutzman H. Reliability and clinical concepts underlying global judgments in demenita: implications for clinical research. Psychopharmacol Bull 1992;28:425-32.

34. Leber P. Slowing the progression of Alzheimer Disease: Methodologic issues. Alzheimer Dis Assoc Disord 1997;11:S10-S21.

35. Sano M, Ernesto C, Thomas RG, et al. A controlled trial of selegiline, alpha-tocopherol, or both as treatment for Alzheimer's disease. The New England Journal of Medicine 1997;336:1216-22.

36. Knapp MJ, Knopman DS, Solomon PR, Pendlebury WW, Davis CS, Gracon SI. A 30-Week Randomized Controlled Trial of High-Dose Tacrine in Patients with Alzheimer's Disease. Journal of American Medical Association 1994;271:985-91.

37. Olin JT, Schneider LS, Doody RS, et al. Clinical evaluation of global change in Alzheimer's disease: Identifying consensus. J Geriatr Psychiatry Neurol 1996;9:176-80.

38. Food and Drug Administration. Peripheral and Central Nervous System Drugs Advisory Committee. Meeting #34. In; March 18, 1993; Rockville, MD; March 18, 1993. p. 1-230.

39. Karlawish JHT, Klocinski J, Merz JF, Clark CM, Asch D. Caregivers' preferences for the treatment of patients with Alzheimer's disease. Neurology 2000;55:1008-14.

40. Canadian Coordinating Office for Health Technology Assessment. Technology Report. Drug Treatments for Alzheimer's Disease. I. A Comparative Analysis of Clinical Trials; 2000.

41. Levinson W, Roach K, Altkorn D, Stern S. Update in general internal medicine. Ann Intern Med 1998;129:212-20.

42. Summers WK. Tacrine in the Treatment of Alzheimer's Disease. Medical Sentinel 2000;5:15-8, 31.

43. Schneider LS. Assessing outcomes in Alzheimer Disease. Alzheimer Dis Assoc Disord 2001;15S:S8-S18.

44. Van Gool WA, Weinstein HC, Scheltens PK, Walstra GJM. Effect of hydroxychloroquine on progression of dementia in early Alzheimer's disease: an 18-month rendomised, double-blind, placebo-controlled study. Lancet 2001;358:455-60.

45. Mohs RC, Doody RS, Morris JC, et al. A 1-year, placebo-controlled preservation of function survival study of donepezil in AD patients. Neurology 2001;57:481-88.

46. Winblad B, Engedal K, Soininen H, et al. A 1-year, randomized, placebo-controlled study of donepezil in patients with mild to moderate AD. Neurology 2001;57:489-95.

47. Marks HM. The Progress of experiment: Science and therapeutic reform in the United States, 1900-1990. Cambridge: Cambridge University Press; 1997.

48. Marks MM. Anatomy of a controversy: The University Group Diabetes Program Study. In: The Progress of Experiment: Science and Therapeutic Reform in the United States, 1900-1990. Cambridge: Cambridge University Press; 1997:197-228.

49. Epstein S. Impure Science: AIDS, activism, and the politics of knowledge. Berkeley: University of California Press; 1996.