



Alzheimer Activism: How To Modernize Clinical Trials?

31 March 2008. On March 13, just north of Washington, DC, leaders of a veritable alphabet soup of Alzheimer disease activist groups met with representatives from the Food and Drug Administration (FDA), clinicians, and industry leaders to discuss ways to improve clinical trials for candidate Alzheimer disease drugs. The workshop focused on how best to measure whether a candidate AD drug is effective. Specifically, the question of the day was what constitutes a clinically meaningful outcome that can support FDA approval for the upcoming generation of new, mechanism-based therapies such as anti-amyloid and neuroprotective drugs, and whether the current FDA requirements should change.

The title of the meeting was “Measures of Clinical Meaningfulness—An AD Ally/FDA Scientific Workshop.” It took place in the same hotel that hosted an expert meeting that gave rise, in the early 1990s, to the current FDA definition of what constitutes a clinically meaningful effect. A sense that this definition needs to evolve pervaded the meeting. This workshop was the first of what is to be a series of discussions between various AD stakeholders and FDA scientists. The workshops aim to engage the agency in a continuing dialog about the changing science of AD, and to reinforce the pressing need for new AD drugs. The goal is to help agency scientists adjust their standards of evaluation for emerging drugs accordingly.

Presentations preceded a long panel discussion. Just below the surface of substantive, genteel exchange, a driving sense of urgency by patient representatives clashed with a cautious—some even said paternalistic—approach by the agency and some doctors. Intense discussion revolved around the question of what standard of proof a biomarker has to meet to win the FDA’s blessing as a surrogate to cut short the length, cost, caregiver burden, and variability of clinical trials. Another flash point was how much risk individual patients versus the agency are willing to accept in drug testing. There was a palpable concern that effective drugs might fail because they are tested too late in the disease process. Furthermore, the discussion highlighted that as more drug candidates than ever before are awaiting human testing, the new bottlenecks in the pipeline now are limitations in diagnostic expertise for early AD, and capacity of clinical trial sites, as well as an insufficient number of patients available for trials. Speakers agreed that this problem calls for future trials to be designed to place less of a burden on caregivers, so that they can fit research participation into their challenging daily lives. As it is, large trials currently compete for the same limited number of patients, and enrollment is slowing down, taking up to 18 months for some trials. Also needed are outreach efforts that appeal to patients’ altruism. Such efforts should educate patients and caregivers about the clinical trials process, ask them to participate for the greater good even if they get randomized to placebo, and offer open-label drug extensions after the blinded period ends.

The workshop was hosted by Accelerate Cure/Treatments for Alzheimer's Disease (ACT-AD). Formed in 2006, this is a coalition of some 50 national organizations representing patients, providers, caregivers, older Americans, researchers, and employers. ACT-AD worked with the Alzheimer's Association and the Alzheimer Study Group (see ARF related news story) to organize the meeting. Other groups represented included the Alzheimer's Drug Discovery Foundation (ADDF), which brings a venture philanthropy approach to translational AD research, the Alzheimer's Foundation of America (AFA), which is primarily concerned with improving care, and the Geoffrey Beene Foundation Alzheimer's Initiative. This meeting report summarizes the scientific presentations that framed the issues, as well as the subsequent discussion.

The View from the Regulator's Perch

What is an effective therapy? Easy as pie—one that changes the life of the patient for the better. Yet as soon as one begins to define this answer, its simplicity disappears, **Bill Thies** of the Alzheimer's Association said in his introductory remarks. This problem dominated the scientific program beginning with **Russell Katz**, who directs the Division of Neurology Products at the FDA's Center for Drug Evaluation Research. Katz laid out the definition for clinical meaningfulness the agency has used since the early 1990s during the time the cholinesterase inhibitor drugs were being tested. At the time, the FDA and AD experts agreed that any AD drug demonstrably needed to do two things: affect a measure of cognition plus have an effect that matters to the patient. The thought was that just because patients score higher on a given memory test scale, this does not automatically mean they feel better. The double requirement, then, was a statistically significant difference on a formal measure of cognition and on a formal measure of clinical improvement. The latter could be a functional measure, or the caregiver's, patient's, or physician's global sense that the patient was better overall.

The FDA continues to insist on the combination of the two measures. Since the early 1990s, most sponsors have come to use the ADAS-cog (a 70-point scale, higher is worse) for the former, and the CIBIC+ (7-point scale, higher is worse) for the latter. Katz noted that once the FDA approved one drug based on results from these scales, other drug companies used them, too. However, the FDA has never mandated the use of any specific batteries, Katz added. Some companies are now shifting to other test packages that they consider more sensitive than ADAS-cog at picking up subtle decline in the earliest stages of AD, for example, the Neuropsychological Test Battery (NTB; see ARF PBT2 story).

Previously, drug sponsors (mostly pharma companies) typically ran three- to six-month placebo-controlled trials with some 200 patients with mild to moderate AD in each group. This gave them enough power to detect very small differences between drug and placebo. In fact, the FDA never defined how large the treatment effects had to be, nor did it require that the patients actually improve relative to their baseline. The agency considered drug effects as small as -1 on ADAS-cog and -0.2 on CIBIC+ to be meaningful, and granted approval when a statistically significant difference between treatments and placebo on two different outcome measures was there, Katz said. In other words, the FDA standard is high in terms of requiring two separate outcomes, but low in the sense that a drug need only nudge these two outcomes a bit.

The agency has been using these same standards for a long time. Katz acknowledged that because the field now is on the verge of a new set of treatments with a different mode of action, the time has come to reconsider them. He offered these alternatives for discussion, ordered by disease stage. Instead of using ADAS-cog, a sponsor could use a test that assesses only a single cognitive function. This is helpful in mild AD when a specific domain fades, such as executive function or episodic memory, while the patient is not broadly impaired yet. Katz insisted that a functional measure was still important to assure the patient actually benefits.

For patients with mild cognitive impairment (MCI), the requirement for some global improvement applies, as well. That issue has not been solved between drug sponsors and the agency. The agency accepts delay to time of AD diagnosis as a clinically meaningful outcome measure. Here, too, the agency does not set a minimum requirement of how long the delay has to be, or how many people reap this benefit, in order to approve.

For asymptomatic (AD scientists would say presymptomatic) patients who have some evidence of impairment, such as subjective memory complaint, or have imaging or CSF biochemistry indicators of AD, a drug-induced change in these measures would be insufficient, Katz said. He said such data couldn't prove that this change will ever translate into improved function, and the agency was unwilling to assume as much. The agency at this point is not prepared to accept a biomarker change as a clinical surrogate. Sponsors are not formally proposing drug trials using only biomarkers to the FDA just yet, Katz said, but added that this is coming. The larger question is what kind of evidence can give the FDA confidence to predict that the patient will get better after a biomarker change. At present, a biomarker change is not a primary outcome. "Approaches that rely on indirect evidence and assumptions are problematic," Katz said. (This point drew critical discussion later on. After the conference, some attendees privately expressed concern that CSF biomarkers were still viewed with skepticism despite a preponderance of evidence. Other attendees said that they expected this view to change as biomarker data from current drug trials come online; see ARF related news story). For normal people at elevated risk, a growing number of scientists want to push treatment back toward secondary and even primary prevention. Katz insisted that the agency was in the dark about what outcome measures it should demand for such trials. Regarding people who have a family history or a strong genetic predisposition, Katz said, dishearteningly: "We are not anywhere near doing those trials yet." (See eFAD trials essays.)

A separate debate revolves around disease modification. The major claim by which the current crop of experimental medicines set themselves apart from approved drugs is that they will slow the underlying disease process, not just provide symptomatic relief. This, too, sounds simple but is difficult to define, much less prove in trials. Uncertainty has reached the point where some clinicians even recommend that industry abandon the distinction altogether, and just try to find a new treatment that works regardless of what words end up on the drug label. The basic problem, as described by many drug company researchers, is the following: the curves of symptomatic and disease-modifying drug effects over time suggest that while the former show up within weeks or a few months after a patient starts taking the drug, a disease-modifying effect may take up to two years

to become measurable. That's because patients do not initially improve on the drug, as they will on a symptomatic drug; they merely decline less steeply than control patients. If the drug effect is weak, say, a 25 percent reduction in decline, it takes a long time for the curves to diverge enough for the trial to pick up the difference. The regulatory take on this situation is skeptical on two accounts. Not only is the FDA unwilling to accept such a change in slope as proof of disease modification, it is also unclear how meaningful such an effect can be to the patient and caregiver if it acts so slowly, Katz said. Discussion about disease modification versus symptomatic drugs, shapes of curves, and timelines led some attendees to conclude that to rise above this muddle, the new crop of drugs simply must show much larger effects than the cholinesterase inhibitor ever did.

The View from the Academic Medical Center

David Knopman of the Mayo Clinic in Rochester, Minnesota, offered a clinical perspective on the challenges of establishing that drug effects matter to patients. Knopman noted that besides the size of drug effects, their limited endurance is a problem—they tend to fade after some months. In general, Knopman added, many clinicians agree with the FDA's dual requirement of a cognitive and a global effect. Practically speaking in drug trials, even small cognitive effects are easy to quantify but impress neither the patients nor their caregivers, whereas global improvement is difficult to measure but meaningful. One trial solved the problem of defining a clinically meaningful result by asking each patient's caregiver to name a valued activity of daily living whose loss would constitute an endpoint, and then measured the time it took patients to decline to this endpoint on donepezil versus placebo (Mohs et al., 2001). This is called survival analysis, and it has promise for broader use in early-stage patients.

For trials in symptomatic patients, Knopman said that in his view, the main open questions are whether a combination of a biomarker and a clinical endpoint might be superior to the current standard, and whether shorter, six-month trials should be sufficient to support FDA approval. Knopman agreed with Katz that finding meaningful change in prevention trials is difficult and expensive. Because no intermediate outcomes are formally accepted yet, the low incidence of AD (1-4 percent newly diagnosed per year in ages 70-90) necessitates large trials that enroll thousands of patients and follow them for years until they develop dementia symptoms. Until such intermediate outcomes are accepted, few primary prevention trials will be undertaken, Knopman said. Thus, open questions remain regarding results from the smattering of such trials to date, including a trial of blood pressure drugs and stroke-induced dementia (Tzourio et al., 2003), and a folate trial that showed cognitive but not clinical improvement (Durga et al., 2007).

Similar questions plague drug trials for mild cognitive impairment (MCI), which lie in between primary prevention trials and trials in full-blown AD patients. As a diagnostic category, MCI is still hotly debated. Some view it as a condition that elevates the risk of AD, while others view it as an artificial construct that contains about 70 percent of people who, with more sensitive measures, can clearly be identified as having very early AD. Put simply, it's a lumpers-versus-splitters argument about how to deal with a disease continuum. In practice, the boundaries between normal and MCI, and MCI and AD, are blurry enough that while one ADCS drug study reported that MCI has worked well as a

diagnostic category (Petersen et al., 2005), other trial sponsors have encountered problems, Knopman noted.

A different clinical perspective came from **Jeffrey Cummings** of the David Geffen School of Medicine of University of California, Los Angeles. Cummings said that the FDA has provided good guidance but is now facing a new era. Advances in understanding the molecular mechanisms of AD have given rise to mechanism-based drugs that should be administered as early as possible, especially given a growing consensus among researchers that the diagnosed phase of AD represents the end phase of a disease process that has been silently occurring in the person for many years before symptoms appeared. This has created new circumstances for drug development and calls for a fresh look at the trial parameters that were developed for the symptomatic drugs of the 1990s, Cummings said. Cummings believes that disease modification is a sensible concept to pursue. He defined disease modification as a drug's ability to slow the underlying AD pathogenetic process and change its clinical course (which cholinesterase inhibitors don't do), not merely the ability of a drug to temporarily delay the disease course (which cholinesterase inhibitors do do). Cummings suggested these talking points for the collective discussion in the field:

Drug sponsors should adopt innovative trial designs such as staggered start and staggered withdrawal (Whitehouse et al., 1998; McDermott et al., 2002) and explore their potential to establish disease modification.

Trials should integrate biomarkers and focus on establishing correlations between markers and clinical outcomes. Specific markers and outcomes could be reasonably matched (e.g., global atrophy to global clinical impression, cingulate atrophy to behavior, hippocampal atrophy to episodic memory).

What is the smallest effect size that matters? Called the minimally clinically important difference (MCID), this concept has received surprisingly little study (Burback et al., 1999; Vellas et al., Lancet Neurology 2008, in press). MCID at present is determined statistically, but really should be determined clinically by the patient/caregiver, physician, and/or regulator. This would raise the bar.

The placebo problem needs a solution. In some recent trials, the placebo group has shown less than the expected rate of decline, and indeed was statistically indistinguishable from the group taking the drug. Part of the reason is that participants in U.S. trials typically receive cholinesterase and/or memantine treatment nowadays. In contrast, the Russian drug Dimebon (see ARF related news story) showed a larger difference between the treatment and placebo groups than is now typical in the U.S., in part because the Russian placebo group did not receive those AD drugs. The challenge of measuring decline in the placebo group is even more acute in very mild AD, because the expected rate of decline is very slow, and so longer trials are needed to show a treatment effect (Gold et al., 2007). Intent-to-treat analysis, where the trial includes data on all randomized people, including those who dropped out, has become standard in clinical drug studies, but it can stack the decks against a drug and end up underestimating its effect. Alternative data

analyses, for example, per protocol observed case (PPOC) analysis, could support temporary marketing approval until larger studies in populations that were not enriched or pre-selected can provide additional safety data.

The more weakly a drug slows progression, the longer trials have to continue before the effect emerges; however, more patients drop out as trials drag on and intent-to-treat analysis then weakens the trial's result. Given these opposing constraints, a 25 percent slowing might represent a workable minimum clinically significant difference, Cummings suggested.

For disease-modifying drugs, finding patients at the earliest possible stage of disease is much more important than it was for symptomatic drugs. To this end, an effort is underway to modernize diagnostic criteria (Dubois et al., 2007). Cummings suggested that a specific early AD symptom—for example, asking the same question over and over again as happens in episodic memory impairment—coupled with a biomarker reading that indicates AD, could be sufficient to include a person in a trial. That person would be diagnosed as having “AD without dementia.” This dual requirement would eliminate from trials the 30 percent of non-AD cases commonly found in MCI populations, who are thought to have diluted the drug effect in prior trials (see Knopman above.)

Both slope analysis (i.e., how does the drug change rate of decline?) and survival analysis (i.e., how quickly does the patient lose a pre-set function or progress to a pre-set clinical dementia rating?) by themselves don't prove that a treatment has really changed the underlying disease process. Cummings agreed with Katz that a clinical measure is important in early-stage patients. He doubted that global clinical measures are appropriate and suggested that behavior offers better clinical outcome information at that stage. Current trials largely ignore psychiatric symptoms of AD, and their inclusion would shift AD drug development toward compounds that also treat this aspect of the disease. For example, trials could measure reduction of aberrant behaviors such as delusions and agitation that were specified at baseline.

In summary, Cummings recommended that the dialogue about meaningful outcome measures for future trials focus on defining measures that truly reflect effects on the underlying disease, on bridging the gap between biomarker and clinical endpoints in trials, and on ways to include earlier-stage patients.

The View from the Trenches

These were the FDA's top-down societal perspectives, and the academic physician-researcher's attempts to imbue clinical trials with the latest science. After that, **Howard Fillit** of the Alzheimer's Drug Discovery Foundation in New York City concluded the presentations with a view from the bottom up. Himself a geriatrician, Fillit asked what clinical meaningfulness means in the daily practice of the family physician or internist, the medical practitioners who care for the majority of AD patients. In essence, for the physician to decide to prescribe an AD drug, the effect of the drug must be robust enough to survive the reality of a 10-minute office visit during which the practitioner assesses a host of simultaneous diseases.

What are the challenges? Typically, a full-time family practitioner sees three or four AD cases per week, and diagnoses few new cases per year, especially of early-stage, younger patients where the incidence is low. Physicians in community practices frequently fail to recognize AD, due to limited or outdated training in dementia. As long as patients manage to maintain their social personality and carry on a short conversation during the exam, their early-stage AD is unlikely to get picked up, Fillit said. Even when it is, the physician may be reluctant to prescribe dementia medication especially for older patients who typically are already taking many other medications and so are at high risk of adverse drug interactions. In addition, prescribing additional drugs may force patients/caregivers to face insurance/Medicare caps for drug reimbursement. Another problem is that, among caregivers and primary care providers, AD drugs are viewed as being less effective than they actually are, Fillit said. Given all this, diagnosing AD frequently is not a top priority for the prescribing physician.

Fillit further noted a regulatory challenge in setting the right standards for clinical meaningfulness. In practice, people's attitudes toward AD vary widely. Some patients and caregivers demand drug treatment to the full extent possible and see value even in small benefits such as a temporary slowdown in decline, while others refuse anything short of a cure. These extremes reflect widely different stances in people's personal views on clinical meaningfulness.

Statistical outcomes are a prerequisite to getting new drugs on the market, but those data help little in getting it to patients, Fillit said. Physicians use neither the ADAS-cog nor the CIBIC scale. What matters most to patients and their caregivers is the patients' ability to function at home. For the physician, the functional benefit of a drug must be straightforward to observe or measure. For example, functional life expectancy is an outcome that offers a straightforward way to explain drug effects to patients (Dodge et al., 2003). The prospect of a year more with one's grandchildren would be meaningful to an early-stage patient, Fillit said. By the same token, disease-modifying therapies at advanced stages of AD may prolong a period of profound disability, he cautioned. Caregiver burden could serve as a meaningful outcome as well, for example, by measuring how much less time the caregiver spends helping the patient with activities of daily living.

Fillit expressed doubt that a drug approved to slow progression by, say, the suggested minimum of 25 percent, would easily take hold in practice. The patient would still get worse on the drug. At the three-, six-, and nine-month visits, the physician would have to persuade the caregiver that it's worth putting up with the cost, side effects, and compliance problems because the patient would have declined a little more without the drug. In that situation, a measurable surrogate marker would help, Fillit said. By analogy, patients stay on statin therapy not because they can feel how the drug slows their atherosclerosis, but because a blood test tells them their cholesterol is down. Similarly, a surrogate marker would be helpful to move an AD treatment into mainstream care, Fillit said.

Patient versus Expert—A Tense Conversation

The workshop concluded with a panel discussion led by **Sid Gilman** of the University of Michigan Medical School in Ann Arbor. The panel brought together Katz, Knopman, Cummings, and Fillit, with **Dale Schenk** of Elan Pharmaceuticals in South San Francisco, Meryl Comer, Bill Bridgewater, a former business executive who has early-onset Alzheimer disease, and his wife Twyla.

Gilman started this conversation by saying that the way toward what he defined as the Holy Grail, i.e., a treatment that truly halts the basic underlying pathophysiology, could be found in innovative trials, such as staggered start trials with survival analysis and strong emphasis on biomarker data. Schenk emphasized that incorporating biomarkers into diagnosis and drug testing is the way forward. He noted that CSF biomarkers have been studied for well over a decade, and more than 300 papers have produced a compelling body of evidence that they work with an accuracy rate as high as that of a good clinician (e.g., Nitsch et al., 1995; Galasko et al., 1998; Sunderland et al., 2003; Hansson et al. 2006; Fagan et al., 2007). That they are still not widely used, nor considered formally validated, is by now an outdated obstacle holding back the field, Schenk said. He added that industry is ready to move a large number of new compounds into trials, and absolutely needs biologic endpoints with accepted metrics to be able to measure drug effects more easily. Katz replied that by law, the FDA is indeed allowed to approve drugs based on surrogates. Even so, the agency requires what he called a “face-valid” outcome, i.e., a measurable clinical improvement, because it is not willing to assume that the surrogate effect alone means the patient will get better. Katz cited historic examples where biomarkers did improve in the expected way, yet the patients did worse (see ARF related biomarker story). Reading between the lines, several attending scientists expressed hope that Katz was holding the party line in public, while preparing to change the agency’s thinking internally. The general sense among scientists at the workshop was that new clinical trials will break the impasse around biomarkers. Historically, imaging studies have not examined patients closely, and detailed clinical studies have included few biomarkers. Drug trials with strong biomarker components effectively force investigators to marry these different outcomes, Schenk said.

The ensuing discussion pitted the patient representatives’ sense of acute urgency against the cautious deliberations of the regulator and the medical establishment. This conversation is growing louder as early-stage patients themselves are increasingly advocating for more national awareness of AD and are pressuring funders, regulators, and drug developers to do much more. For example, the Bridgewaters serve on the FDA’s Alzheimer Disease Advisory Committee and speak at events around the country (see also Q&A with patient-advocate Richard Bozanich). At the ACT-AD workshop, Bill Bridgewater urged that the FDA fast-track candidate AD drugs in the way HIV drugs previously were. (The FDA last year fast-tracked AZD-103; see also ARF related news story.) His wife Twyla urged that cutting-edge drugs be more available to caregivers, who can become exhausted, poor, and depressed from the burden of many years of care. Meryl Comer, president of the Geoffrey Beene Foundation Alzheimer’s Initiative, described the toll that 12 years of caring for her husband, and now also her elderly mother, have taken. That toll is physical (i.e., sheer exhaustion from 24/7 care, with never

enough help, for a heavier, stronger, frustrated spouse), professional (Comer left her job as a top broadcast journalist), financial, and emotional. “My definition of clinical meaningfulness is simple: A hand that lifts itself to feed, a mind that does not see the caregiver as an enemy,” Comer said.

Much of the discussion revolved around the contentious question of how much risk is acceptable. Comer asked regulators and drug developers to take on more risk both in how drug trials are conducted, and in terms of side effects. “We will happily manage side effects; that is what families do,” she said. Fillit noted that a recent survey reflected a growing willingness among patients and caregivers to accept risk from side effects as serious as stroke or even death. But accused of lax oversight by the media and Congress every time an approved drug makes the headlines for unexpected side effects, the FDA so far has shown little appetite for accepting more risk in either trial design or adverse event considerations for what is considered a particularly vulnerable patient population. Comer rejected this stance as paternalistic. She charged that if the FDA were to cling to old standards for clinical effectiveness, the agency would allow AD to become the untreated epidemic of the current generation. Bridgewater noted that he would tolerate a high degree of risk, knowing full well he has a terminal illness and remembering how it claimed his father. Cummings cautioned that not all risk is the same. A side effect that can be treated and resolved is one thing; a stroke that leaves the patient permanently worse, quite another. “I don't want that for my patients, because AD patients up to a certain level of impairment continue to enjoy their life,” he said.

For his part, Katz defended the FDA's position by pointing to the many drugs with safety risks that are on the market with black-box labels. The agency has turned down very few AD drugs because they were too risky, he said. He assured the audience that side effects per se would not derail an otherwise effective drug. “If we had a drug that prevented AD and had a meaningful effect, we might tolerate a fair amount of risk,” Katz said. The discussion sounded as if the FDA is more prepared to tolerate the risk of side effects than the risk that a biomarker-driven drug might end up not working. The patient representatives countered that the FDA should leave more room for educated patients and their doctors to make decisions about which drug to choose. In closing, Gilman called on the various stakeholders to see themselves as a team, where everyone has a different role in the goal of making better medicines for AD.—Gabrielle Strobel.