Alzheimer's & Dementia: Translational Research & Clinical Interventions 3 (2017) 402-409

Review Article

Phase II clinical trials of anti–amyloid β antibodies: When is enough, enough?

Michael Gold*

Global Product Development, Neuroscience, PPD, Morrisville, NC, USA

Abstract Efforts to develop new therapies to combat Alzheimer's disease suffer from extraordinarily high failure rates that make it difficult to justify continued investment in the field. Although there are a number of plausible explanations for this extremely high attrition rate, one of the explanations that has received little attention is the lack of compelling data from Phase II studies for compounds that have been pushed into Phase III trials and then have failed. An analysis of publicly available data from the Phase II studies for bapineuzumab and solanezumab indicates that neither compound produced compelling evidence of drug-like behavior that would justify their progression into pivotal trials. The published data suggest that sponsors took decisions to move these compounds into Phase III on the basis of vastly limited data that were rife with type I error and probably driven by commercial concerns. The continued push to move compounds that are not likely to succeed in later stage clinical trials threatens to erode trust in the clinical research enterprise making it much harder to properly test truly promising compounds.

© 2017 The Author. Published by Elsevier Inc. on behalf of the Alzheimer's Association. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Keywords: Alzheimer's disease; Clinical trials; Pharmacology; Phase II; Phase III; Monoclonal antibodies; Amyloid β; Investment

1. Introduction

The quest for effective disease modifying agents against Alzheimer's disease (AD) represents a complex scientific challenge and a critical public health goal [1]. AD represents the most common primary neurodegenerative disorder, and there is a general consensus that the cost of caring for the increasing number of persons affected by AD will create serious public health problems [2]. At the same time, there is recognition that the etiology of AD is multifactorial and that simple reductionist hypotheses are not likely to yield tractable solutions [3].

One of the most popular and dominant hypotheses in the area of AD is the amyloid hypothesis [4,5], which posits that either the overproduction and/or the underclearance of amyloid β (Aβ) is a proximate cause of AD. This hypothesis—bolstered by the evidence from patients carrying mutations affecting the APP, PS-1, or PS-2 genes [6] and evidence that Aβ deposition may play a role in initiating neurodegeneration—has driven a massive investment from industry, government, and academia to find ways to attack the amyloid cascade. Research into clearance mechanisms of Aβ suggested that either active [7] or passive immunization [8] approaches could be potentially useful to reduce the amount of Aβ in the brain and thereby prevent delay or even possibly reverse the cognitive and functional decline that defines AD.

Despite the vast amount of information gathered for more than the last 30 years on the pathogenesis of AD, the track record for compounds targeting the amyloid pathway is

Disclaimer: This manuscript was written by Michael Gold, MD in his personal capacity. The opinions expressed in this article are the author's own and do not reflect the view of PPD.

Funding: None.

*Corresponding author. Tel.: +1-919-457-6679; Fax: 1-919-633-1298.
E-mail address: michael.gold.md@gmail.com

http://dx.doi.org/10.1016/j.trci.2017.04.005

2352-8737/ © 2017 The Author. Published by Elsevier Inc. on behalf of the Alzheimer's Association. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

abysmal with no single Phase III trial reporting a positive result on a primary outcome [9]. The extensive number of studies (clinical and preclinical) related to Aβ notwithstanding fundamental questions [10] critical to the success of clinical trials such as the actual toxicity of the peptide remains unanswered or poorly understood [11].

One of the currently popular explanations for these failures is that treating AD patients when they are demented is too late and that, to modify the natural course of the disease, one needs to intervene when persons are cognitively intact or very mildly affected, although it is not known just how early one would need to start treatment [12]. As a consequence, a number of clinical trials are now testing a variety of Aβ–based prevention studies in people who are cognitively normal but at a higher risk of developing AD by virtue of apolipoprotein E (*APOE*) status or by having a high Aβ load detected using positron emission tomography (PET) imaging [13].

There are multiple alternative explanations for these clinical trial failures including (1) the possibility that the amyloid hypothesis is wrong [14]; (2) the possibility that treatments aimed at a single pathologic process will be ineffective [15]; and (3) the possibility that the amyloid hypothesis may be correct, but that the compounds that have been taken into clinic are ineffective [16] and do not represent a true test of the amyloid hypothesis in symptomatic patients.

The first alternative explanation leads to a falsifiable hypothesis that treating patients who carry mutations known to cause early onset AD before neurodegeneration sets in should be able to delay or prevent the onset of dementia. In this respect, the fact that prevention studies are being conducted in populations carrying APP or PS mutations is reassuring [17]. The second alternative explanation also leads to testable and falsifiable hypotheses about the potential efficacy of combination therapies based on known pathophysiological aspects of AD. The third explanation leads to at least two potential root causes: (1) the preclinical models currently used to test compounds are not appropriate [18] and are systematically biased toward "false positive" results; and/or (2) compounds are being pushed into pivotal trials despite a lack of robust signals of "drug-like behavior" at earlier stages of development. This last potential root cause for several recent and prominent clinical trial failures is the focus of this article.

What are the harms of pushing compounds into Phase III prematurely? From a societal perspective, the loss of trust and credibility between study sponsors, investigators, and potential subjects (particularly ethnic minorities [19]) and their treating physicians [20] in clinical trials stands out as the most serious harm. When sponsors move compounds into Phase III trials prematurely and encounter issues with lack of efficacy or safety issues, the trust and goodwill of future potential subjects is squandered, making the already challenging recruitment and retention of subjects [21] much harder for subsequent trials [22]. This is particularly true when clinical trials are complex, require invasive procedures and repeated imaging procedures, and demand participation for several years [23].

If the harms from pushing compounds into Phase III prematurely can pose an existentialist threat to the clinical research enterprise, one is bound to ask what incentives can lead to such behaviors, particularly because the vast majority of people working to develop new drugs for AD patients are smart, passionate, and well intentioned.

From a financial perspective, the revenue forecasts from being a disease-modifying agent approved for AD may provide the rationale for proceeding into Phase III despite risks that would not be acceptable in other indications. With peak-year sales for bapineuzumab estimated to be ~$5–10B [24] and solanezumab projected to earn ~$5B [25] within a few years of approval, the protection from generics afforded by biological agents and the sunk costs associated with development and manufacturing, even a 1% probability of success may be acceptable from a financial perspective. However, the cost of failure is not borne solely by the sponsor who decides to take a compound into Phase III prematurely. When large Phase III trials in AD fail, the financial community's reaction is not limited to the compound or sponsor in question, but signals a greater overall lack of confidence in AD clinical trials, outcome measures, biomarkers, so forth. This "collateral" damage was evident after the failure of the Phase III trials of bapineuzumab and the exit from neuroscience research of several large pharmaceutical companies noting that research in AD was "just too risky" [26]. A more recent example of this phenomenon was the drop in the stock price of companies working on alternative mechanisms when the solanezumab Expedition 3 trial was announced. One would have expected that companies working on alternative mechanisms would have been spared, but that was not the case.

There is at least one other source of misaligned incentives when it comes to the decision to progress a compound into Phase III, namely the "academic-industrial" complex [27]. Historically, clinical research at academic medical centers was funded primarily through grants and supplemented by clinical revenue. However, in the face of decreased compensation for routine clinical care and flat federal funding with its emphasis on basic research, some clinical researchers are increasingly reliant on industry-sponsored funding [28]. In many cases, the experts advising sponsors on clinical development strategies and the design of clinical trials are the same authorities who will participate in the trial as paid investigators and serve as paid consultants and promotional speakers, and the same experts who will benefit professionally from the multitude of publications that are generated by clinical trial programs. A feed-forward cycle has been set up in which the interdependence [29] between pharmaceutical companies, academic institutions, and investigators [30] likely rewards decisions to move compounds into late stage development even if the data do not warrant it.

My hope is that exploring one of the potential reasons why the bapineuzumab and solanezumab Phase III

programs failed in an admittedly critical and provocative way will foster a candid and productive dialogue resulting in improvements in drug development. This article focuses on bapineuzumab and solanezumab because they share a mechanism of action (passive clearance of amyloid β, although targeting different portions of the peptide) and they both completed large Phase III trials that failed to detect clinically relevant treatment effects [31,32]. I reviewed the published Phase II data for both compounds with an eye toward the robustness of signals of efficacy and contrast that with the timing of the decision to start Phase III development. Fundamental elements of successful drug development, such as evidence of target engagement, dose or exposure-response relationships, signals of efficacy, and convergence of data, were sought out in each of the Phase II programs.

2. Bapineuzumab

The primary Phase II study for bapineuzumab reported by Salloway et al. in 2009 [33] is described as "initially designed and powered to evaluate the safety of bapineuzumab" and was then "amended to evaluate efficacy as the primary objective based partly on the preliminary results from the Phase I study." The study was initially powered to have a >80% probability of detecting adverse events occurring with a rate of at least 5% within each dose cohort. This leaves one wondering what was really accomplished by amending the study to test for efficacy when neither the details of the amendment nor the parameters related to efficacy (effect size, power, alpha, variance) are specified. Interestingly enough, the authors comment that "the urgency of delivering an effective treatment to patients with AD argued for making efficacy the primary outcome." If that were truly the case, then one would expect the authors to report the efficacy data from the Phase I study that provided the rationale for the sample size selection in the Phase II trial.

The sponsor chose the Alzheimer's Disease Assessment Scale - cognitive subscale (ADAS-Cog)-12 as a coprimary outcome measure despite known limitations of the battery and the availability of more sensitive instruments that could have provided cleaner signals of efficacy [34,35] given the very small sample sizes in each dose cohort. One can speculate that the decision to use the ADAS-Cog/Disability Assessment for Dementia (DAD) as coprimary outcome measures was driven by the desire to get point estimates for effect size in anticipation of using them in registrational studies. The prespecified statistical analysis "compared bapineuzumab to placebo within the 0.5, 1.0, and 2.0 mg/kg cohorts based on change from baseline to week 78" using a repeated measures model in the modified Intent To Treat (mITT) population. There is no discussion around type I error or control for multiplicity in these three dose arms, but there is text to the effect that the 0.15 mg/kg dose arm was included in the exploratory analyses and there was no correction for multiplicity in that dose arm. As reported by the authors, no significant differences between bapineuzumab and placebo on either of the coprimary outcome measures at the 0.5, 1.0, and 2.0 mg/kg dose were detected. Examination of Table 2 reveals that there is no dose-response relationship for either outcome measure. The largest observed (and nominally significant, but exploratory) effect on the ADAS-Cog was seen at the 0.15 mg/kg dose, but not corroborated by data from the DAD at that dose. The authors then argue that, given the discrepancy between observed and modeled data and the limited power given the small sample size in each dose cohort, additional exploratory analyses in which dose arms were pooled were appropriate. Although it is unclear why the discrepancy between observed and modeled data justifies the pooling of dose groups, these pooled analyses failed to detect statistically significant effects on the ADAS-Cog, neuropsychological test battery (NTB), and DAD. In terms of target engagement, analyses of data from subjects who agreed to undergo lumbar punctures ($n = 20$ on bapineuzumab and $n = 15$ on placebo) failed to detect any effect on $A\beta_{1-42}$ or total tau, but detected a trend ($P = .056$) favoring bapineuzumab on tau_{p181} levels. In addition, analyses of volumetric magnetic resonance imaging (MRI) data did not detect a treatment effect in the pooled dose groups. Post hoc subgroup analyses based on *APOE* status yielded nominally significant results on the ADAS-Cog, NTB, Mini–Mental State Examination (MMSE) and Clinical Dementia Rating Sum of Boxes (CDR-SB), but not on the DAD. It should be noted that cognitive batteries tend to have a degree of colinearity so these "positive results" are not totally unexpected. There are no results reported on dose-response relationship within the *APOE* noncarriers. There are at least 30 *P* values reported or implied in the article so using a nominal significance of .05, there is a 79% probability of seeing at least one *P* value $\leq .05$ by random chance. In terms of criteria that are generally looked at to progress a compound into Phase III, data on target engagement are not supportive, data in terms of dose or exposure-response are not supportive, and signals of potentially clinically relevant benefits are also lacking. Elan/Wyeth announced the start of the Phase III program for bapineuzumab on May 21, 2007 on the basis of "the seriousness of the disease and the totality of what the companies have learned from their immunotherapy programs, including a scheduled interim look at data from an ongoing Phase II study, which remains blinded" [36]. If the results reported in 2009 represent the totality of the data from this Phase II study and are found to be lacking, it is hard to comprehend what could have possibly been so compelling at an interim look, particularly when the single most "effective" dose (0.15 mg/kg) was tested relatively late in the conduct of the study.

A second study by Rinne et al. [37] was published in 2010 and reported the outcome of a Phase II study designed to investigate the effects of treatment with bapineuzumab on measures of cerebral Aβ deposition using ^{11}C-Pittsburgh compound B (^{11}C-PIB) PET imaging and cerebrospinal fluid (CSF) biomarkers, MRI parameters and cognition for more than a period

of 78 weeks. Subjects (~ 10/arm) were randomized to receive infusions of 0.5, 1.0, or 2.0 mg/kg every 13 weeks in a 7:3 active to placebo ratio. The planned sample size was expected to provide >95% power to detect a change from baseline in mean ^{11}C-PIB retention ≥0.25 between the pooled placebo and pooled bapineuzumab arms with an $\alpha = 0.05$. The primary outcome was the mean ^{11}C-PIB retention cerebral to cerebellar retention ratio across six predefined cortical areas. The planned analysis took into account screening ^{11}C-PiB, baseline MMSE (as a categorical value [high vs. low]), visit week, and an interaction term between treatment and visit week and was carried out on the mITT population (subjects who received any amount of investigational drug and had a baseline and at least one valid PET scan postbaseline).

The authors note that because of observed imbalances between the treatment and placebo groups at baseline on the NTB, CDR-SB, and ^{11}C-PIB, post hoc analyses were carried out that adjusted for these imbalances. Recruitment into the 2.0 mg/kg dose was truncated because of reports of vasogenic edema from the study described previously. Fifty-three subjects were screened; 28 (20B:8P) were randomized, 26 (19B:7P) were in the mITT population, and 20 (15B:5P) had ^{11}C-PiB scans on week 78.

Inspection of Fig. 2 suggests that the active and placebo-treated populations were on differing trajectories with increasing ^{11}C-PIB signal in the placebo group and decreasing signal in the active group such that a statistically significant difference was detected at week 78. However, Fig. 3 suggests that the increase in the ^{11}C-PIB signal in the placebo group may have been driven by a single outlier. Using a program designed to extract numerical data from plots (http://arohatgi.info/WebPlotDigitizer), data from Fig. 3 were extracted and compared with the published data. The published baseline values for the mean ^{11}C-PiB retention ratios for the placebo and bapineuzumab groups were 1.89 and 2.06, respectively, and the extracted values were 1.88 and 2.06, respectively, indicating that the extraction program generated valid data. The mean (standard deviation [SD]) value for all seven placebo patients was 2.029 (0.0173). When the single outlier (value = 2.4) was removed, the mean (SD) dropped to 1.968 (0.069), which is lower than the extracted mean (SD) for the bapineuzumab group of 1.978 (0.206). It is not possible to determine what effect that single outlier may have had on the primary efficacy analysis, but given how small the placebo group is and the fact that the separation between the two groups is driven by the larger increase in signal in the placebo group (0.15) compared with the smaller drop in signal in the bapineuzumab group (−0.09), the effect of even a single outlier cannot be underestimated.

The authors note that no effects were seen on any clinical, fluorodeoxyglucose (FDG)-PET, MRI, or CSF end points after adjustment for the baseline imbalances noted previously. It should also be noted that for many of the ^{11}C-PiB measurements, the adjusted analyses resulted in upper limits of 95% confidence intervals that were very close to zero (−0.002 to −0.03), again suggesting that statistical significance may have been driven by a single data point. The authors note that the differences between treatment groups were similar in each of the doses tested, yet at the same time they caution that the sequential recruitment of small cohorts to ascending doses limited the capacity to detect dose-response effects. Because the study was powered to detect an effect on the pooled groups, one has to wonder why the study was designed with dose-escalation rather than a single (presumably effective) dose. The fact that the highest dose (2.0 mg/kg) could not be tested in a full cohort because of safety issues and that the most effective dose in the previous Phase II study (0.15 mg/kg) was not included in this study leaves many questions unanswered as to how to interpret this study.

3. Solanezumab

Siemers et al. [38] reported the outcome of a placebo-controlled, ascending single-dose Phase I/II study in patients with mild to moderately severe probable AD. Cohorts of five subjects were enrolled and randomized (4:1) to active or placebo at doses of 0.5, 1.5, 4.0, and 10.0 mg/kg. Subjects were followed through 365 days for safety purposes. As this was the first study in humans, the principal objectives of the study were to characterize the safety, pharmacokinetics, and pharmacodynamics of single doses of Sola. In terms of pharmacodynamics, changes from predose (day −2) to postdose (day 21) were sought in plasma and CSF measures of various species of Aβ and the ADAS-Cog 11 item battery. Analysis of CSF Aβ indicated a trend ($P = .05$) for the slope of the dose versus change from predose to day 21 in CSF $A\beta_{1–42}$ to differ from 0, but no such trend was detected for $A\beta_{1–40}$. No signal of effect was detected on the ADAS-Cog 11. This study suggests that there is target engagement (at least in the periphery) and possibly in the central nervous system (CNS). There is no evidence of a dose-response relationship or a signal on clinical measures, but this study was clearly not designed or powered to answer such questions.

In 2012, Farlow et al. [39] reported the outcome of a placebo-controlled, parallel-group study that investigated the effects of treatment with three different doses of solanezumab infused weekly for 12 weeks in subjects with mild to moderately severe probable AD. Subjects were randomized in a 4:1 ratio to active (100 mg every 4 weeks, 100 mg once weekly, 400 mg every 4 weeks, and 400 mg once weekly) or placebo with the primary objective of the study being to assess the safety and tolerability of multiple doses. Secondary objectives included assessments of plasma and CSF pharmacodynamics for dose selection and assessment of cognitive effects of short-term administration using the ADAS-Cog 11- and 14-item versions. There is no rationale for the proposed sample size, but the cohort structure and size are typical for an exploratory multiple-ascending dose study designed to detect relatively frequent adverse events. The analysis of changes in Aβ CSF concentrations and ADAS-Cog scores was carried out using upper-tailed, one-sided hypothesis tests with α set at 0.1.

Analytical validation of plasma and CSF Aβ assays indicates that the assays were stable, of high quality, and reliable. Increases in plasma $A\beta_{1–40}$ were dose-proportionate but were less clearly dose-proportionate for $A\beta_{1–42}$. Total CSF $A\beta_{1–40}$ and $A\beta_{1–42}$ was increased in the highest three doses compared with placebo or to the 100 mg every 4 weeks dose; however, there was no difference in the effect between the 100 mg once weekly and 400 mg every 4 weeks doses both of which had effects that were statistically superior to the 100 mg every 4 weeks dose, but numerically closer to that dose than to the 400 mg once weekly dose. In contrast, analyses of the unbound fraction of $A\beta_{1–40}$ and $A\beta_{1–42}$ in the CSF failed to detect any dose-response relationship although the three highest doses resulted in statistically significantly higher $A\beta_{1–42}$ concentrations compared with placebo. No signals of efficacy were detected on either the 11- or 14-item version of the ADAS-Cog at any dose or at any time point and nor was there any suggestion of a dose/exposure-response relationship. Although the authors point out that in a study of this duration, expectations regarding effects on cognition should be low, they also point out that -effects were seen on memory tests after single doses in a transgenic mouse model leaving the reader to determine just how high or low to set their expectations. With the availability of more sensitive cognitive batteries and specifically given the small sample sizes in this study, one is again hard-pressed to explain the choice of the ADAS-Cog as the sole cognitive outcome scale in this study.

As predicted from the pharmacologic properties of solanezumab, dose-proportionate increases in antibody-bound plasma Aβ were detected. The increase in CSF total Aβ suggests that some solanezumab may have been able to cross the blood-brain barrier and suggests target engagement in the CNS. The observed increase in free $A\beta_{1–42}$ concentrations was attributed by the authors to plaque solubilization and re-establishment of equilibrium in the CSF after sequestration of Aβ in the plasma [40]. Although this is a plausible hypothesis, the authors seem to gloss over the fact that this particular species of Aβ is more neurotoxic than the 1–40 species [41] and that sustained exposure to higher levels of the 1–42 species may not be a good thing for AD patients. The authors conclude that "on the basis of this study and the changes seen in CSF Aβ, several multinational Phase III trials were started."

The fact that no effects were seen on other markers of neurodegeneration (tau or brain volume) and that there is no discussion about what criteria needed to be met for progression into Phase III leaves one wondering just how low a threshold needed to be crossed to justify the start of Phase III trials. Lilly announced the start of the Expedition studies in May 2009 suggesting that a decision to move into Phase III trials was made sometime in 2008, which is consistent with the completion of the Phase II study in May 2008.

4. Summary

Bapineuzumab and solanezumab represent two compounds that specifically target AD disease modification by increasing the clearance of Aβ. In the case of bapineuzumab, the data from the two available Phase II studies fail to provide evidence that the compound had the necessary properties to be successful in Phase III. The lack of compelling Phase II data may be one of the reasons why the original sponsor (Elan) sought out partners for the development of this compound. It is entirely plausible that when the cost of development was spread out among two or three partners, the risk of failure became tolerable in the context of a potentially massive payoff if it had been successful. In the other case, we have a single Phase II study of solanezumab, which suggests there was target engagement in the CNS, therapeutic doses were being tested, and there was a hint of dose-response relationship with a single but relevant CSF biomarker. Neither compound was able to establish any signal of efficacy on measures of cognition.

After the failure of the expedition and Expedition 2 studies, another explanation surfaced (the studies failed because ~20%–30% of subjects did not have pathologic burdens of Aβ). This explanation is somewhat suspected because subjects without adequate pathology would have been randomized to both placebo and active treatment arms and sample size calculations would have accounted for this heterogeneity and because a substantial proportion of patients who do not have pathologic amyloid burden still convert to AD dementia [42]. However, the argument that modest treatment effect sizes could be masked may be more compelling.

The recent report of the results of the Expedition 3 study [43] merits some discussion given the comments from some experts in the field that, despite the fact that it failed to detect a clinically or statistically significant difference on its primary end point, it should be viewed as confirmation of the amyloid hypothesis. On first principles, it should be clear that studies do not confirm hypotheses, they either provide evidence to support a hypothesis or they refute a hypothesis. It is also naive to claim that a single study can confirm a given hypothesis based on our troubling inability to replicate results in biomedical research. One can speculate about what would have happened had Lilly opted to use the CDR-SB or the Alzheimer's Disease Cooperative Study-Activities of Daily Living as the primary outcome measure given that statistically significant but clinically irrelevant differences were detected between treatment groups on these outcomes. This is why the principle of requiring cognitive and global coprimaries espoused by Paul Leber (former Director of the Division of Neuropharmacological Drug Products at the FDA from 1981 to 1999) was developed: to prevent the gaming of the system by using very large sample sizes to detect statistically significant but clinically irrelevant treatment effects. It is interesting to note that the ADAS-Cog failed to detect an effect whereas the CDR-SB did, raising the question of whether treatment effects in noncognitive domains or cognitive domains not included in the ADAS-Cog may have accounted for effects on the CDR-SB and

Alzheimer's Disease Cooperative Study-Activities of Daily Living. It is hard to understand why statistically significant differences were detected on the MMSE, but not on the ADAS-Cog-14. Given the size of Expedition 3 ($n \sim 2100$), the duration of treatment (80 weeks), the tight inclusion/exclusion criteria (mild severity and verification of amyloid pathology), the results, including the failure to detect a significant reduction in amyloid burden on PET imaging, suggest minuscule treatment effects at best (Cohen's $d = 0.07$, 0.11, and 0.1 for the ADAS-Cog-14, ADL, and CDR-SB, respectively [44]) and statistical noise at worst. There are differing opinions about whether Expedition 3 produced any data supportive of the amyloid hypothesis, but it certainly did not provide data to refute it. Had there been profound effects on CSF or PET measures of amyloid or brain atrophy that failed to translate into delays in progression of cognitive decline, it seems we would probably agree that the amyloid hypothesis had been properly tested. So it seems we are still in the hunt for a compound that can provide a real test of the hypothesis.

The failure of these two compounds also has been used as the rationale for studies in asymptomatic, at-risk subjects. Although there may be merit to treating at-risk subjects as early as possible to increase the odds of modifying the course of the disease using the lowest-possible dose to maximize safety and reduce costs, the fact that these two compounds failed does *not* falsify or disprove [45] the hypothesis that treating symptomatic subjects is rational because, as I hope has been made clear, the Phase II data for both these compounds were far from compelling.

So where do we go from here and when is enough, enough? Some suggestions for improving the quality of Phase II programs already have been promulgated by the Alzheimer's Association Research Roundtable [46] and sponsors are strongly urged to consider these. Implementing independent replication of pilot or Phase II studies [47], documentation of AD pathology for targeted therapy, prospectively defined go-no-go criteria, stricter control over type I error, and post hoc analyses [48] only can help to strengthen the case for a compound that truly merits going into Phase III. In addition, proposals to increase the validity of published data have been proposed [49]. Some common sense suggestions would be to (1) conduct independent Phase II studies to provide replication of a signal of effect with one study focusing on CSF biomarkers and another on imaging biomarkers; (2) keep single ascending dose/multiple ascending dose studies simple and focused on dose selection based on pharmacokinetics/pharmacodynamics, safety, and tolerability; (3) use of Bayesian or permutation tests to get a sense of the robustness of the data rather than reliance on P values (particularly with small sample sizes); (4) use adaptive testing to avoid floor/ceiling effects on cognitive batteries and ensuring the batteries cover all relevant domains; (5) test more than one dose in the Phase II studies to explore dose/exposure-response relationships as a guide to dose selection; and (6) use non-transgenic animal models in which Aβ deposition takes place over a long period of time (e.g., aged dogs, microcebus, and guinea pigs).

An example of the kind of study that should be conducted before entering into Phase III is the one reported by Sevigny et al. [50]. This study uses a placebo-controlled, parallel-group design to test a range of doses of aducanumab on Aβ burden measured by PET imaging. The sample sizes in each arm are substantial ($n \sim 30$–40) and provided >90% power to detect at least a 1 SD difference in amyloid reduction relative to baseline comparing each group to placebo with an $\alpha = 0.05$. Subjects were screened to make sure they had pathologic burdens of amyloid in support of a diagnosis of mild or prodromal AD. The study included the MMSE and more sensitive measures such as the NTB and the Free and Cued Selective Reminding Test as exploratory cognitive outcome measures. Although the results presented in this publication are based on interim data, the exploration of a broad range of doses and the dose-proportionate response in terms of Aβ reduction reported in the article are strong markers for target engagement and bode well for the future development of this compound. That said, Phase III trials were launched before the completion of the PRIME study and there is no explanation as to why treatment effects were not detected on sensitive measures of cognition. Interestingly enough, beneficial effects were reported in subjects completing 12 months of an open-label extension on the MMSE and CDR-SB (the same instruments showing an effect in Expedition 3).

The leaders of pharmaceutical and biotech companies need to understand and appreciate the consequences of pushing compounds into Phase III prematurely not only to their own organizations, but for the entire clinical research enterprise. Reward and incentive systems within organizations need to push toward healthier behaviors (e.g., bonuses paid on informative studies rather than on the number of compounds transitioned from Phase II to Phase III or the start of a Phase III program). Stringent and transparent controls are needed to prevent financial conflicts of interest between clinical experts and sponsors developing programs. In addition, the clinical research communities (academic, private, and commercial) need to revisit their dependence on clinical trial volume and focus on supporting those clinical trials that are based on robust, high-quality science to recruit high-quality subjects quickly and properly test the most promising hypotheses.

The failure to address these issues presents us with the potential for a "tragedy of the commons" in which sponsors continue to push weak compounds into Phase III whose inevitable failure will erode the willingness of patients and their families to participate in trials, the willingness of health care providers to refer patients to clinical trials, and the willingness of investors to fund novel treatments.

RESEARCH IN CONTEXT

1. Systematic review: Phase II clinical trials for bapineuzumab and solanezumab were identified via online searches using PubMed and references from the manuscripts reporting the pivotal trials for both compounds.
2. Interpretation: The phase II data for both compounds indicate that "drug-like" characteristics were lacking for both compounds and that progression of these compounds into pivotal trials was probably not scientifically justified. The total amount of phase II data (numbers of subjects, duration of exposure, doses, so forth) for both these compounds is very limited and the interpretation of the phase II studies is hampered by uncontrolled type I error. The conclusion that antiamyloid antibodies will not work in symptomatic patients is not supported.
3. Future directions: The use of novel trial designs, independent replication of studies, and prespecified go-no-go criteria are encouraged to increase confidence in the value of phase II data and to reduce the risk of failure.

References

[1] Winblad B, Amouyel P, Andrieu S, Ballard C, Brayne C, Brodaty H, et al. Defeating Alzheimer's disease and other dementias: a priority for European science and society. Lancet Neurol 2016;15:455–532.

[2] Prince M, Bryce R, Albanese E, Wimo A, Ribeiro W, Ferri CP. The global prevalence of dementia: a systematic review and metaanalysis. Alzheimers Dement 2013;9:63–75.e2.

[3] Jellinger KA. Alzheimer's disease: a challenge for modern neuropathobiology. Acta Neuropathol 2009;118:1–3.

[4] Hardy J, Selkoe DJ. The amyloid hypothesis of Alzheimer's disease: progress and problems on the road to therapeutics. Science 2002; 297:353–6.

[5] Selkoe D. The molecular pathology of Alzheimers disease. Neuron 1991;6:487–98.

[6] Bateman RJ, Xiong C, Benziger TLS, Fagan AM, Goate A, Fox NC, et al. Clinical and biomarker changes in dominantly inherited Alzheimer's disease. N Engl J Med 2012;367:795–804.

[7] Kokjohn TA, Roher AE. Antibody responses, amyloid-beta peptide remnants and clinical effects of AN-1792 immunization in patients with AD in an interrupted trial. CNS Neurol Disord Drug Targets 2009;8:88–97.

[8] DeMattos RB, Lu J, Tang Y, Racke MM, DeLong CA, Tzaferis JA, et al. A plaque-specific antibody clears existing β-amyloid plaques in Alzheimer's disease mice. Neuron 2012;76:908–20.

[9] Cummings JL, Morstorf T, Zhong K. Alzheimer's disease drug-development pipeline: few candidates, frequent failures. Alzheimers Res Ther 2014;6:37.

[10] Armstrong R. A critical analysis of the "amyloid cascade hypothesis". Folia Neuropathol 2014;3:211–25.

[11] Hardy J. The amyloid hypothesis for Alzheimer's disease: a critical reappraisal. J Neurochem 2009;110:1129–34.

[12] Pike KE, Savage G, Villemagne VL, Ng S, Moss SA, Maruff P, et al. β-Amyloid imaging and memory in non-demented individuals: evidence for preclinical Alzheimer's disease. Brain 2007;130:2837–44.

[13] Sperling RA, Rentz DM, Johnson KA, Karlawish J, Donohue M, Salmon DP, et al. The A4 study: stopping AD before symptoms begin? Sci Transl Med 2014;6:228fs13.

[14] Giacobini E, Gold G. Alzheimer disease therapy-moving from amyloid-β to tau. Nat Rev Neurol 2013;9:677–86.

[15] Stephenson D, Perry D, Bens C, Bain LJ, Berry D, Krams M, et al. Charting a path toward combination therapy for Alzheimer's disease. Expert Rev Neurother 2015;15:107–13.

[16] De Strooper B. Lessons from a Failed γ-Secretase Alzheimer Trial. Cell 2014;159:721–6.

[17] Reiman EM, Langbaum JBS, Fleisher AS, Caselli RJ, Chen K, Ayutanont N, et al. Alzheimer's prevention initiative: a plan to accelerate the evaluation of presymptomatic treatments. Adv Alzheimer Dis 2011;2(Suppl 3):589–97.

[18] Van Dam D, De Deyn PP. Drug discovery in dementia: the role of rodent models. Nat Rev Drug Discov 2006;5:956–70.

[19] Roberson N. Clinical trial participation. Cancer 1994;74:2687–91.

[20] Jenkins V, Fallowfield L. Reasons for accepting or declining to participate in randomized clinical trials for cancer therapy. Br J Cancer 2000;82:1783–8.

[21] Fargo KN, Carrillo MC, Weiner MW, Potter WZ, Khachaturian Z. The crisis in recruitment for clinical trials in Alzheimer's and dementia: an action plan for solutions. Alzheimers Dement 2016;12:1113–5.

[22] Gauthier S, Albert M, Fox NC, Goedert M, Kivipelto M, Mestre-Ferrandiz J, et al. Why has therapy development for dementia failed in the last two decades? Alzheimers Dement 2016;12:60–4.

[23] Grill JD, Karlawish J. Addressing the challenges to successful recruitment and retention in Alzheimer's disease clinical trials. Alzheimers Res Ther 2010;2:34.

[24] Carroll J. Bapineuzumab—15 top blockbuster contenders. FierceBiotech. Available at: http://www.fiercebiotech.com/special-report/bapineuzumab-15-top-blockbuster-contenders. Accessed December 10, 2016.

[25] Taylor L. Alzheimer's drug market "set to triple by 2022." PharmaTimes. 2013. Available at: http://www.pharmatimes.com/news/alzheimers_drug_market_set_to_triple_by_2022_1005934. Accessed December 10, 2016.

[26] Stovall S. R&D cuts curb brain-drug pipeline. Wall St J. 2011. Available at: http://www.wsj.com/articles/SB10001424052748704474804576222463927753954. Accessed December 11, 2016.

[27] Relman A. The new medical-industrial complex. N Engl J Med 1980; 303:963–70.

[28] Moses H, Matheson DHM, Cairns-Smith S, George BP, Palisch C, Dorsey ER. The anatomy of medical research. JAMA 2015; 313:174–89.

[29] Moses H 3rd, Thier SO, Matheson DHM. Why have academic medical centers survived? JAMA 2005;293:1495–500.

[30] Zeller C. The pharma-biotech complex and interconnected regional innovation arenas. Urban Stud 2010;47:2867–94.

[31] Salloway S, Sperling RA, Fox NC, Blennow K, Klunk W, Raskind M, et al. Two phase 3 trials of bapineuzumab in mild-to-moderate Alzheimer's disease. N Engl J Med 2014;370:322–33.

[32] Doody RS, Thomas RG, Farlow M, Iwatsubo T, Vellas B, Joffe S, et al. Phase 3 trials of solanezumab for mild-to-moderate Alzheimer's disease. N Engl J Med 2014;370:311–21.

[33] Salloway S, Sperling RA, Gilman S, Fox NC, Blennow K, Raskind M, et al. A phase 2 multiple ascending dose trial of bapineuzumab in mild to moderate Alzheimer disease. Neurology 2009;73:2061–70.

[34] Maruff P, Thomas E, Cysique L, Brew B, Collie A, Snyder P, et al. Validity of the CogState brief battery: relationship to standardized tests and sensitivity to cognitive impairment in mild traumatic brain injury, schizophrenia, and AIDS dementia complex. Arch Clin Neuropsychol 2009;24:165–78.

[35] Duff K, Clark JDH, O'Bryant SE, Mold JW, Schiffer RB, Sutker PB. Utility of the RBANS in detecting cognitive impairment associated with Alzheimer's disease: sensitivity, specificity, and positive and negative predictive powers. Arch Clin Neuropsychol 2008; 23:603–12.

[36] Elan. Elan and Wyeth to initiate phase 3 clinical trial of bapineuzumab (AAB-001). Available at: https://www.alzconnected.org/WorkArea/threadeddisc/print_thread.aspx?id=1436&g=posts&t=36590. Accessed December 13, 2016.

[37] Rinne J, Brooks D, Rossor M, Fox N. 11C-PiB PET assessment of change in fibrillar amyloid-β load in patients with Alzheimer's disease treated with bapineuzumab: a phase 2, double-blind, placebo-controlled, ascending-dose study. Lancet 2010;4422:1–10.

[38] Siemers ER, Friedrich S, Dean RA, Gonzales CR, Farlow MR, Paul SM, et al. Safety and changes in plasma and cerebrospinal fluid amyloid beta after a single administration of an amyloid beta monoclonal antibody in subjects with Alzheimer disease. Clin Neuropharmacol 2010;33:67–73.

[39] Farlow M, Arnold SE, vanDyck CH, Aisen PS, Snider BJ, Porsteinsson AP, et al. Safety and biomarker effects of solanezumab in patients with Alzheimer's disease. Alzheimers Dement 2012; 8:261–71.

[40] Demattos R, Bales K, Cummins D, Paul S, Holtzman D. Brain to plasma amyloid-b efflux: a measure of brain amyloid burden in a mouse model of Alzheimer's disease. Science 2002;295:2264–7.

[41] Mucke L, Selkoe D. Neurotoxicity of amyloid β-protein: synaptic and network dysfunction. Cold Spring Harb Perspect Med 2012;2:1–17.

[42] Caroli A, Prestia A, Galluzzi S, Ferrari C, van der Flier WM, Ossenkoppele R, et al. Mild cognitive impairment with suspected non-amyloid pathology (SNAP): prediction of progression. Neurology 2015;84:508–15.

[43] Lilly announces detailed results of solanezumab Phase 3 EXPEDITION 3 study at the Clinical Trials on Alzheimer's Disease (CTAD) 2016 Meeting. 2016. Available at: https://investor.lilly.com/releasedetail.cfm?ReleaseID=1003490. Accessed January 8, 2017.

[44] Schneider L. A calculated perspective on solanezumab's outcomes. Alzforum. 2016. Available at: http://www.alzforum.org/news/conference-coverage/ctad-solanezumab-seen-nudge-ad-ever-so-slightly#comment-22866. Accessed January 8, 2017.

[45] Popper K. The logic of scientific discovery. London: Routlege Classics; 2005.

[46] Greenberg BD, Carrillo MC, Ryan JM, Gold M, Gallagher K, Grundman M, et al. Improving Alzheimer's disease phase II clinical trials. Alzheimers Dement 2013;9:39–49.

[47] Loscalzo J. Pilot trials in clinical research: of what value are they? Circulation 2009;119:1694–6.

[48] Pocock SJ, Assmann SE, Enos LE, Kasten LE. Subgroup analysis, covariate adjustment and baseline comparisons in clinical trial reporting: current practice and problems. Stat Med 2002;21:2917–30.

[49] Ioannidis JPA. How to make more published research true. PLoS Med 2014;11:e1001747.

[50] Sevigny J, Chiao P, Bussiere T, Weinreb PH, Williams L, Maier M, et al. The antibody aducanumab reduces Aβ plaques in Alzheimer's disease. Nature 2016;537:50–6.