

## Aducanumab Trials EMERGE But Don't ENGAGE

L.S. Schneider

Corresponding Author: Lon S. Schneider, MD, Keck School of Medicine of USC, Los Angeles, USA, [lschneid@usc.edu](mailto:lschneid@usc.edu)

If at first you don't succeed, try, try (and try?) again to get that darned manuscript published just the way you want it (1). Which is what 24 authors did after rejections or 'revise and resubmits' from JAMA, probably from the New England Journal of Medicine before JAMA, if not another journal afterwards (Just guessing based on an Axios report and the timing of the JAMA submission <https://www.axios.com/biogen-jama-aduhelm-clinical-trial-results-publish-fc7c2876-a684-4bfc-8462-4165f57d735a.html>).

The Emerge (302) and Engage (301) manuscript appears to have been withheld over two years because the Biogen and academic authors would not respond or comply with reviewers and editors. The manuscript could have been reviewed by 6 to 12 reviewers before JPAD got its turn. This published version, unfortunately, shows scant evidence of having been critically reviewed (1). There are defensive tones here and there, but little evidence of deliberation or uncertainty in the presentation, conduct and outcomes of the trials. (It's perhaps notable, that their structured abstract remains in the unique style that JAMA requires for submitted manuscripts and that is reformatted when the revised manuscript is published (<https://jamanetwork.com/journals/jama/pages/instructions-for-authors#SecAbstractsforReportsofOriginalData>). Clearly, the authors wanted complete control of their message and not to have to acknowledge the substantial limitations of their trials, outcomes, and inferences they make.

The message they want to sell is: High dose aducanumab met its "prespecified" primary and secondary endpoints; showed an association between reduction of biomarkers of "underlying disease pathology" and slowing of clinical decline; and has a "clinically meaningful" effect. Frankly, this is so wrong on many levels.

### Methods

The trials were not executed or analyzed as planned and were fraught with methodological challenges and unforced errors. Amendments 3 and 4 (PV4), instituted after a substantial proportion of patients had been randomized, allowed treatment to be restarted after brain edema or hemorrhage (ARIA), and allowed

APOE4 carriers to receive the highest, 10 mg/kg, dose where previously they were restricted to 6 mg/kg. The amendments contributed to the substantial, practical unblinding of the treatments as these procedural switches and a 35 to 40% rate of ARIA, mainly affecting the 70% of patients who were APOE4 carriers, couldn't be kept from patients and many research staff. Moreover, exercising a protocol-specified option to increase sample size altered expectations, prolonged the recruitment period just as it was winding down, and further increased the high risk for bias. It is unfortunate, but not unusual, that the methods and treatment dose for Alzheimer phase 3 trials are still being worked out during the trials rather than having been set beforehand. Although the authors will tell us that they could detect no bias in either study it's important to consider that bias cuts both ways.

### Futility Analysis

Futility decisions rely on a planned, conditional probability estimate that a trial has a certain likelihood to be statistically significant as expected in the protocol if it were to finish. When futility is declared as per the protocol it is the same as a final analysis. The data monitoring committee has reviewed the futility analysis, checked the conditional probability estimates, and basically did an interim and final efficacy analysis. Here, as well, the DMC discussed results with Biogen who shared in the decision to stop (2). After futility is determined subsequent analyses are post hoc, subset analyses, and not "prespecified" as Biogen would have it. Moreover, multiplicity, interim efficacy analyses for stopping early, and adjustments for Type I error correction come into play.

Biogen later argued that the assumptions underlying their futility analysis were violated. This is like saying, 'Sorry, we did the wrong analysis, but never mind.' According to Biogen these assumptions were, first, that the treatment effect in the two trials would be similar, and second, that there is a constant effect throughout the trials, meaning that the later enrolled patients would show the same effect as earlier enrolled patients.

Some of this is inherently illogical because if the treatment effects are different between identically designed trials, then the trials don't replicate. One does not confirm or support the other; the validity of both

may be questioned. Facing discordant results while maintaining optimism, Biogen should have immediately started a third phase 3 trial in the Spring of 2019 when they first learned their outcomes. FDA fully understands the need for replication and confirmation as it first argued without apparent irony to its advisory committee in November 2020 that the negative Engage trial did not contradict Emerge and wasn't needed for regular approval because a phase 1b study could fill the role of a confirmatory trial. Subsequently, as a condition for accelerated approval FDA required that a third, similar phase 3 trial be done as a post-marketing requirement.

The second assumption, that there is a constant effect throughout the trials, is one made in all trials that do not explicitly treat participants randomized early versus later differently in the pooled result. Put simply, in trial designs we expect the first patient to be like the last. Yet, it is often true that later participants may differ from earlier recruited participants in that the pool of easily available participants becomes depleted or because trial centers recruit at differing rates. If this is the explanation for why the futility analysis failed, then the treatment outcomes should only apply to this early subgroup while adjusting the p value to correct for type I error because of the multiple analyses. However, since the trials were stopped about halfway through this would be difficult to calculate, be roughly equivalent to an interim analysis to stop early for efficacy and require a much more stringent boundary for statistical significance than a nominal  $P = .05$ . Indeed, Biogen's statistical analysis plan called for an O'Brien-Fleming stopping boundary for an interim analysis for efficacy (2) (FDA biostatistics report, page 20) and thus the critical P value would be much lower than .01.

In any event, the authors want us to forget the futility analysis, ignore the Engage trial, accept the unusual changes to the Emerge placebo group as ordinary, discount the functional unblinding and other risks for bias, and simply accept Emerge as a positive trial with "clinically meaningful" outcomes. We cannot.

## Prespecified Analyses

The statement that the study "followed prespecified statistical analyses" is faulty unless Biogen had a plan for what to do after stopping for futility as knowledge of the futility results makes all these subset analyses post hoc. Most of the analyses are exploratory, not corrected for multiplicity or false discovery, and yet are misrepresented as prespecified. The FDA biostatistics report noted that Biogen did not follow their own planned sequential testing procedures as intended and assuming that the trials had been completed and not truncated. The Type I error could be as high as .0975, i.e., 10%, if one could pick and choose between the high and low doses in the studies (2) (FDA biostatistics report, page 18). Certainly, it is greater than the  $P = .012$  reported for the high dose in Emerge and mischaracterized by Biogen's representatives

as robust and strongly significant. The sequential testing procedures required that the high-dose group be tested against placebo first, then the low-dose group. If both are significant, then the secondary outcomes are tested in a specific order. This did not happen. The critical point is that only the high-dose group of one trial was nominally significant, i.e.,  $CDR-sb = 0.39$  (95% CI -0.69 to -0.09),  $P=0.012$ , and the other high- and low-doses were not. Under these conditions none of the secondary outcomes could be considered significant, and it is misleading for the authors to say, "data from EMERGE demonstrated a statistically significant change across all four primary and secondary clinical endpoints".

## Correlating Clinical Measures with Biomarkers

Correlating clinical change with biomarker change is a big issue for FDA accelerated approval guidelines as they do not require evidence of clinical benefit and which aducanumab lacks. Rather, accelerated approval requires mainly the demonstration of an effect on a surrogate endpoint that is "reasonably likely" to predict clinical benefit (3, 4). From the FDA's point of view correlating plaque reduction with clinical ratings, even in negative studies or by pooling databases, helps to meet this "reasonably likely" standard.

Aducanumab's effect on reducing plaques, the effect of fibrils on neurons, and consequently reducing p-tau expression is incontrovertible. This is the biology of the matter and what got aducanumab from phase 1b to phase 3 trials. Remarkably, the authors chose to emphasize the very large, dose effect of aducanumab on amyloid-PET, based only on subsets of the protocol completers who opted for week 78 scans, certainly not an as randomized sample. They highlighted this as the main figure (Figure 2) of their publication (1) while relegating the depiction of the primary and secondary clinical outcomes to the online supplement.

In an apparent effort to defend accelerated approval and establish amyloid-PET as a surrogate clinical endpoint, Biogen authors retrospectively correlated change in amyloid-PET as the predictor variable of change in CDR-sb score (5). At the individual patient level, they reported adjusted Spearman correlations for the pooled high and low dose groups in Emerge and Engage of 0.19 and -0.06, respectively. Perhaps it is not surprising that they showed the pooled coefficient values because the correlation coefficients for the high dose patients in Emerge was less than for the low dose patients, 0.13 vs. 0.21, suggesting, on face, that any relationship of plaque lowering with clinical improvement is with the low dose patients. From the FDA biostatistics report (2) (Table 11, page 56) one can only surmise that they reported Spearman over Pearson coefficients and pooled the high and low treatment groups from Emerge, because that gave the best

coefficient number with the lowest P value out of 36 exploratory correlations.

The dose-group level correlations were interesting in that the correlation was in the predicted direction only when the high-dose group from Engage was left out (Supplemental Data Fig. 4b). In Engage, high dose treatment was associated with about a +0.5-point worsening on the CDR-sb compared to placebo and yet with a substantial reduction in plaques. While in Emerge high dose treatment was associated with about a -0.5-point improvement and with a similar reduction in plaques. In other words, high dose aducanumab was about as cognitively impairing in Engage as it was beneficial in Emerge, and in both trials showing similar, substantial reductions in plaque. To be sure a 0.5 CDR-sb difference is a small effect and not clinically meaningful (6) although it was the effect used to statistically power the two trials.

Despite the trivial and conflicting relationships, the Biogen authors claim in their summary that, "EMERGE is the first phase 3 trial to demonstrate an association between reduction of biomarkers of AD pathology and a statistically significant slowing of clinical decline, supporting the possibility that removal of A $\beta$  from the brain ... may be associated with a clinical benefit in patients with early AD"

This is a stunningly remarkable, false claim without any foundation. To say that these tiny post hoc correlations chosen among many, based on non-randomized, convenience data from only those who finished the trial and agreed to a second amyloid-PET scan have decipherable meaning, let alone is evidence that plaque reduction is a surrogate marker for clinical outcome is about as true as saying the earth is flat.

## The Placebo

If one were to look for the one thing that can account for all the observed differences between the half-completed trials, then it is the placebo group of Emerge showing a 1.74-point, worsening CDR-sb change compared to a smaller 1.56-point change in Engage, the difference representing nearly half the week 78 mean difference between high-dose aducanumab and placebo of -0.39. The other treatment groups were very similar between the trials. The increased placebo group worsening in Emerge could be explained by chance and the greater placebo progression after PV4 occurring only in Emerge.

## Conclusion

A proper understanding of what Biogen presents in this JPAD paper and FDA's Office of Biostatistics' report reveals the lack of evidence for efficacy of aducanumab and explains why the European Medicines Agency rejected it. By contrast FDA's accelerated approval was based on an unvalidated surrogate biomarker and not on

its regular, standard, low bar for substantial evidence in clinical trials (7). The FDA does not view clinical benefit as necessary for marketing approval on the accelerated pathway. Indeed, the demonstration of clinical benefit would have earned aducanumab regular approval. All the FDA must do for accelerated approval is to view a biomarker endpoint as "reasonably likely to predict clinical benefit." "Reasonably likely," however, is essentially a guess or inference and is hardly compelling. All accelerated approval means in the context of Alzheimer pathology is, "We approved aducanumab because it reduces plaques."

This should not sit well with patients, families, physicians, or insurers as there are no apparent clinical benefits or improved health outcomes with aducanumab that can outweigh its harms. Biogen and the FDA dropped an unfinished, undertested, potentially unsafe product, without evidence of benefit, on a vulnerable American public, as if to say, "We're done here, you deal with it".

As a practical matter, time, advancing clinical science, and imminent amyloid fibril antibody trials results will put aducanumab outcomes in perspective. Lecanemab and gantenerumab phase 3 trials results, due to read out in the last quarter of 2022, will give answers that will make the uncertainties of aducanumab moot. If one or the other meets the "substantial evidence" standard, then that treatment will get regular approval. If neither is positive, then we would face the untenable circumstance of three amyloid fibril antibodies with six negative or truncated Phase 2 and 3 trials between them, all having received marketing approval because they reduce plaque or phosphorylated tau, and all lacking clinical benefit. These two antibodies and donanemab, which would have its phase 3 results in Spring 2023, would still have a future, however, in three 'preclinical AD' (meaning people with a positive amyloid biomarker with or without a positive tau marker and without cognitive impairment) prevention trials that would continue for at least the next 5 years. Testing facets of the amyloid cascade hypothesis will grind on.

Meanwhile, be prepared for a flood of tortured, selected, and highly crafted post hoc analyses to purport that the outcomes in Emerge are 'clinically meaningful,' or that there are real correlations between clinical change and biomarker change that means the biomarker is a surrogate clinical marker.

Finally, it's good that Biogen will post aducanumab individual patient data on Vivli.org. Although not fully open and public, and Biogen has a say in who can get access, Vivli provides a chance for others to examine the trials and learn from them.

*Conflict of interest:* Dr. Schneider reports grants and personal fees from Eli Lilly and Roche/Genentech; personal fees from Boehringer Ingelheim, Neurim, Ltd, Cognition Therapeutics, Takeda, vTv, Samus, Immunobrain Checkpoint, Cortexyme, AC Immune, Otsuka, GW Research, Novo Nordisk and Vivli; grants from Eisai, Biogen, Novartis, Biohaven, and Washington University/NIA DIANTU from outside of and within 2 years of this work.

## References

1. Budd Haeberlein S, Aisen PS, Barkhof F, et al. Two Randomized Phase 3 Studies of Aducanumab in Early Alzheimer's Disease. *J Prev Alz Dis* 2022;2(9):197-210; 2022/03/18 2022;10.14283/jpad.2022.30
2. Food and Drug Administration Office of Biostatistics. Statistical Review and Evaluation, NDA/BLA#: 761178 (aducanumab), July 7, 2020, finalized May 11, 2021, [https://www.accessdata.fda.gov/drugsatfda\\_docs/nda/2021/761178Orig1s000StatR\\_Redacted.pdf](https://www.accessdata.fda.gov/drugsatfda_docs/nda/2021/761178Orig1s000StatR_Redacted.pdf)
3. Food and Drug Administration. Expedited Programs for Serious Conditions Drugs and Biologics. 2014. <https://www.fda.gov/files/drugs/published/Expedited-Programs-for-Serious-Conditions-Drugs-and-Biologics.pdf>
4. Dunn B, Stein P, Temple R, Cavazzoni P. An Appropriate Use of Accelerated Approval — Aducanumab for Alzheimer's Disease. *New England Journal of Medicine* 2021; 385(9): 856-7.
5. Thambisetty M, Howard R, Glymour MM, Schneider Lon S. Alzheimer's drugs: Does reducing amyloid work? *Science* 2021; 374(6567): 544-5.
6. Liu KY, Schneider LS, Howard R. The need to show minimum clinically important differences in Alzheimer's disease trials. *The Lancet Psychiatry* 2021; 8(11): 1013-6.
7. Food and Drug Administration. Demonstrating Substantial Evidence of Effectiveness for Human Drug and Biological Products Guidance for Industry (Draft Guidance). 2019. <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/demonstrating-substantial-evidence-effectiveness-human-drug-and-biological-products> (accessed March 23, 2022).

How to cite this article: L.S. Schneider. Aducanumab Trials EMERGE But Don't ENGAGE. *J Prev Alz Dis* 2022;2(9):193-196; <http://dx.doi.org/10.14283/jpad.2022.37>